EXTREME BIASES IN PRIME NUMBER RACES WITH MANY CONTESTANTS

KEVIN FORD, ADAM J HARPER, AND YOUNESS LAMZOURI

Abstract. We continue to investigate the race between prime numbers in many residue classes modulo q, assuming the standard conjectures GRH and LI.

We show that provided $n/\log q \to \infty$ as $q \to \infty$, we can find n competitor classes modulo q so that the corresponding n-way prime number race is extremely biased. This improves on the previous range $n \geqslant \varphi(q)^{\epsilon}$, and (together with an existing result of Harper and Lamzouri) establishes that the transition from all n-way races being asymptotically unbiased, to biased races existing, occurs when $n = (\log q)^{1+o(1)}$.

The proofs involve finding biases in certain auxiliary races that are easier to analyse than a full n-way race. An important ingredient is a quantitative, moderate deviation, multi-dimensional Gaussian approximation theorem, which we prove using a Lindeberg type method.

1. Introduction

Let $q \geqslant 3$ and $2 \leqslant n \leqslant \varphi(q)$ be integers, (where the Euler function $\varphi(q)$ denotes the number of residue classes mod q that are coprime to q), and let $\mathcal{A}_n(q)$ be the set of ordered n-tuples $(a_1, a_2, \ldots, a_n)$ of distinct residue classes that are coprime to q. In this paper we are interested in the "Shanks–Rényi prime number race", which is the following problem: if we let $\pi(x; q, a)$ denote the number of primes $p \leqslant x$ with $p \equiv a \bmod q$, is it true that for any $(a_1, a_2, \ldots, a_n) \in \mathcal{A}_n(q)$, we will have the ordering

$$\pi(x; q, a_1) > \pi(x; q, a_2) > \cdots > \pi(x; q, a_n) \tag{1.1}$$

for infinitely many integers x? There is now an extensive body of work investigating different aspects of this question, and the reader may consult the expository papers of Granville and Martin [7], Ford and Konyagin [5], and Martin and Scarfy [12] for fuller discussions.

Date: August 22, 2018.

Kevin Ford was supported by National Science Foundation grant DMS-1501982. Adam Harper was supported, for part of this research, by a research fellowship at Jesus College, Cambridge. Youness Lamzouri is partially supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. In addition, part of this work was carried out at MSRI, Berkeley during the Spring semester of 2017, supported in part by NSF grant DMS 1440140; and part of the writing up was supported by funding (for Ford and Harper) from the Simons Foundation and the Centre de Recherches Mathématiques, Montréal, through the Simons-CRM scholar-in-residence program.

Assuming the Generalized Riemann hypothesis GRH and the Linear Independence hypothesis LI (the assumption that the nonnegative imaginary parts of the nontrivial zeros of Dirichlet L-functions attached to primitive characters are linearly independent over $\mathbb{Q}$), Rubinstein and Sarnak [14] proved that the answer to this question is always Yes. More strongly, they proved that for any $(a_1, \ldots, a_n) \in \mathcal{A}_n(q)$, the set of real numbers $x \geqslant 2$ such that (1.1) holds has a positive logarithmic density, which we shall denote by $\delta(q; a_1, \ldots, a_n)$. Recall here that the logarithmic density of a subset $S \subseteq \mathbb{R}$ is defined as

$$\lim_{x\to\infty} \frac{1}{\log x} \int_{t \in S \cap [2,x]} \frac{dt}{t},$$

provided the limit exists. This density can be regarded as the "probability" that for each $1 \leqslant j \leqslant n$, the player a_j is at the j-th position in the prime race. As we shall discuss, a probabilistic perspective turns out to be very helpful in this problem.

Next it is natural to ask whether all orderings of the $\pi(x; q, a_i)$ occur with approximately the same logarithmic density, in other words whether $\delta(q; a_1, \ldots, a_n) \approx 1/n!$. For small q, the widely known phenomenon of *Chebyshev's bias* implies that $\delta(q; a_1, a_2)$ can be significantly different from $1/2$, if one of the a_i is a quadratic residue and the other a non-residue mod q. For example, Theorem 1.11 of Fiorilli and Martin [4] implies that $\delta(24; 5, 1) \approx 0.99999$, assuming GRH and LI. On the other hand, Rubinstein and Sarnak [14] showed (assuming GRH and LI) that for any fixed n,

$$\lim_{q\to\infty} \max_{(a_1,\ldots,a_n)\in\mathcal{A}_n(q)} \left| \delta(q; a_1, \ldots, a_n) - \frac{1}{n!} \right| = 0, \tag{1.2}$$

in other words any biases dissolve when $q \to \infty$. Different behaviour is possible if one races "teams" of many residue classes combined against one another (e.g. all the quadratic residues mod q against all the non-residues mod q), as explored in Fiorilli's paper [3]. Feuerverger and Martin [2] raised the question of having a uniform version of Rubinstein and Sarnak's statement, in which $n \to \infty$ as $q \to \infty$. And they asked whether for n sufficiently large in terms of q the asymptotic formula $\delta(q; a_1, \ldots, a_n) \sim 1/n!$ might become false. Ford and Lamzouri (unpublished) formulated the following conjecture.

Conjecture 1.1 (Ford and Lamzouri). *Let $\varepsilon > 0$ be small and q be sufficiently large.*

(1) (Uniformity for small n) If $2 \leqslant n \leqslant (\log q)^{1-\varepsilon}$, then uniformly for all n-tuples $(a_1, \ldots, a_n) \in \mathcal{A}_n(q)$ we have $\delta(q; a_1, \ldots, a_n) \sim 1/n!$ as $q \to \infty$.

(2) (Biases for large n) If $(\log q)^{1+\varepsilon} \leqslant n \leqslant \varphi(q)$, then there exist n-tuples $(a_1, \ldots, a_n) \in \mathcal{A}_n(q)$ and $(b_1, \ldots, b_n) \in \mathcal{A}_n(q)$ for which $n! \cdot \delta(q; a_1, \ldots, a_n) \to 0$ and $n! \cdot \delta(q; b_1, \ldots, b_n) \to \infty$ as $q \to \infty$.

The first part of this conjecture is now known to hold (assuming GRH and LI) in a slightly stronger form, as Harper and Lamzouri [8] proved that, uniformly for all $2 \leqslant n \leqslant \log q/(\log\log q)^4$ and all n-tuples $(a_1, \ldots, a_n) \in \mathcal{A}_n(q)$, we have

$$\delta(q; a_1, \ldots, a_n) = \frac{1}{n!}\left(1 + O\left(\frac{n(\log n)^4}{\log q}\right)\right).$$

This improved on an earlier result of Lamzouri [10], where the asymptotic $\delta(q; a_1, \ldots, a_n) \sim 1/n!$ was established in the range $n = o(\sqrt{\log q})$, assuming GRH and LI.

Regarding the second part of the conjecture, Harper and Lamzouri [8] proved (assuming GRH and LI) that for any $\varepsilon > 0$ and every $\varphi(q)^\varepsilon \leqslant n \leqslant \varphi(q)$, there exists an n-tuple $(a_1, \ldots, a_n) \in \mathcal{A}_n(q)$ such that

$$\delta(q; a_1, \ldots, a_n) < \big(1 - c_\varepsilon\big)\frac{1}{n!},$$

where $c_\varepsilon > 0$ depends only on ε. This was the first result on n-way prime number races where the biases do not dissolve when $q \to \infty$, but it is clearly far from the full statement in part (2) of Ford and Lamzouri's conjecture. In particular, we note that the bias $1 - c_\epsilon$ is always less than 1 (whereas the conjecture asserts that biases towards both small and large values should be possible), and always close to 1 (whereas multipliers that tend to zero or to infinity with q should be possible). Our goal in this paper is to revisit this issue.

We shall prove the following result.

Theorem 1.2. *Assume GRH and LI. There exists a large absolute constant C such that the following is true. Provided n is sufficiently large and $n \leqslant \varphi(q)$, there exist an n-tuple $(a_1, \cdots, a_n) \in \mathcal{A}_n(q)$ such that*

$$\delta(q; a_1, \ldots, a_n) \leqslant \exp\left(-\frac{\min\{n, \varphi(q)^{1/50}\}}{C\log q}\right)\frac{1}{n!},$$

and there exist distinct reduced residues $b_1, \cdots, b_n$ modulo q such that

$$\delta(q; b_1, \ldots, b_n) \geqslant \exp\left(\frac{\min\{n, \varphi(q)^{1/50}\}}{C\log q}\right)\frac{1}{n!}.$$

Notice this fully establishes part (2) of Ford and Lamzouri's conjecture, as soon as $n/\log q \to \infty$. Furthermore, as n becomes larger the relative biases become quantitatively very extreme, and for $n \leqslant \log q$ the bias $\exp\left(\pm\frac{n}{C\log q}\right) = 1 \pm \Theta\left(\frac{n}{\log q}\right)$ roughly matches the factor $\left(1 + O\left(\frac{n(\log n)^4}{\log q}\right)\right)$ in Harper and Lamzouri's [8] uniformity result. (For large fixed n, one can also think of this as clarifying the dependence on n in Theorem A of Lamzouri [11].)

Remark. The assumptions of GRH and LI imply that the summands in the explicit formulae for $\pi(x;q,a)$ behave like independent random variables (see [14] for details), and one can use this property to derive a useful explicit formula for our logarithmic densities $\delta(q;a_1,\dots,a_n)$. See section 2.2, below. If GRH or LI were false, these explicit formulas could have very different behaviour. For example, by [6, Theorem 5.1], for any q, $2\leqslant n\leqslant \varphi(q)$ and $(a_1,\dots,a_n)\in\mathcal{A}_n(q)$, there is a hypothetical configuration of zeros of the Dirichlet L-functions modulo q off the critical line (i.e., violating GRH) which, if they exist, imply that at most $n(n-1)$ of the $n!$ orderings of the functions $\pi(x;q,a_i)_{1\leqslant i\leqslant n}$ occur when x is large enough. In particular, $\delta(q;a_{\sigma(1)},\dots,a_{\sigma(n)})=0$ for most permutations σ, and likewise there exists a permutation such that the upper density of the set

$$\{x:\pi(x;q,a_{\sigma(1)})\geqslant\cdots\geqslant\pi(x;q,a_{\sigma(n)})\}$$

is at least $\frac{1}{n(n-1)}$. Both of these hypothetical conclusions would constitute much larger biases than the bounds given in Theorem 1.2.

We shall actually obtain Theorem 1.2 as a straightforward corollary of another ordering result. For any integers $1\leqslant k\leqslant n\leqslant\varphi(q)$, let us define $\delta_k^\flat(q;a_1,\dots,a_n)$ to be the logarithmic density of the set of real numbers $x\geqslant 2$ such that

$$\pi(x;q,a_1)>\pi(x;q,a_2)>\cdots>\pi(x;q,a_k)>\max_{k+1\leqslant j\leqslant n}\pi(x;q,a_j). \tag{1.3}$$

If everything were uniform, we would expect that $\delta_k^\flat(q;a_1,\dots,a_n)\approx(n-k)!/n!$. Let us also define $\delta_{2k}^\sharp(q;a_1,\dots,a_n)$ to be the logarithmic density of the set of real numbers $x\geqslant 2$ such that

$$\begin{aligned}&\pi(x;q,a_1)>\pi(x;q,a_3)>\cdots>\pi(x;q,a_{2k-1})>\max_{2k+1\leqslant j\leqslant n}\pi(x;q,a_j)\\ &\quad>\min_{2k+1\leqslant j\leqslant n}\pi(x;q,a_j)>\pi(x;q,a_{2k})>\cdots>\pi(x;q,a_4)>\pi(x;q,a_2).\end{aligned}\tag{1.4}$$

Again, if everything were uniform we would expect that $\delta_{2k}^\sharp(q;a_1,\dots,a_n)\approx(n-2k)!/n!$.

Harper and Lamzouri [8] proved the uniformity result that, assuming GRH and LI,

$$\delta_k^\flat(q;a_1,\dots,a_n)=\frac{(n-k)!}{n!}\left(1+O\left(k(\log k)^6\frac{\log n}{\log q}+\frac{1}{n\log^{1/10}q}\right)\right)$$

whenever $k(\log k)^{10}\leqslant(\log q)/\log n$. They also proved a non-uniformity result: for any fixed[1] $k\geqslant 2$, fixed $\varepsilon>0$, and any $\varphi(q)^\varepsilon\leqslant n<\varphi(q)^{1/41}$, there exists an n-tuple $(a_1,\dots,a_n)\in\mathcal{A}_n(q)$

[1]We remark that the case $k=1$ is not excluded from the non-uniformity result for technical reasons, but because it is genuinely different. Harper and Lamzouri [8] showed that $\delta_1^\flat(q;a_1,\dots,a_n)\sim 1/n$ provided $n=o(\varphi(q)^{1/32})$, when $q\to\infty$. Notice that Theorem 1.3, below, deals with $\delta_{2k}^\flat(q;a_1,\dots,a_n)$ and $\delta_{2k}^\sharp(q;a_1,\dots,a_n)$, so the case of $\delta_1^\flat(q;a_1,\dots,a_n)$ is again excluded.

such that
$$\delta_k^\flat(q;a_1,\dots,a_n) < \big(1-c_\varepsilon\big)\frac{(n-k)!}{n!}.$$

Here we establish a significantly improved non-uniformity result.

Theorem 1.3. *Assume GRH and LI. There exists a large absolute constant A such that the following is true. Suppose q is large, and let $1 \leqslant k \leqslant n/A \leqslant \varphi(q)^{1/50}$. Then there exists a tuple $(a_1,\cdots,a_n) \in \mathcal{A}_n(q)$ such that both*

$$\delta_{2k}^\flat(q;a_1,\dots,a_n) \leqslant \exp\left(-\frac{k\log(n/k)}{2\log q}\right)\frac{(n-2k)!}{n!} \tag{1.5}$$

and

$$\delta_{2k}^\sharp(q;a_1,\dots,a_n) \geqslant \exp\left(\frac{k\log(n/k)}{2\log q}\right)\frac{(n-2k)!}{n!}. \tag{1.6}$$

Proof of Theorem 1.2, assuming Theorem 1.3. Suppose first that $n \leqslant \varphi(q)^{1/50}$. Take $k = \lceil n/2A\rceil$, (which indeed satisfies $1 \leqslant k \leqslant n/A$ since we assume in Theorem 1.2 that n is large), and note that by Theorem 1.3 we have
$$\delta_{2k}^\flat(q;a_1,\dots,a_n) \leqslant \exp\left(-\frac{n\log(2A)}{4A\log q}\right)\frac{(n-2k)!}{n!}.$$
On the other hand, since the logarithmic density of the set of real numbers $x \geqslant 2$ for which $\pi(x;q,a) = \pi(x;q,b)$ is 0 (which follows from equation (2.2) below), we get
$$\delta_{2k}^\flat(q;a_1,\dots,a_n) = \sum_{\sigma\in S_{n-2k}} \delta(q;a_1,a_2,...,a_{2k},a_{\sigma(2k+1)},\dots,a_{\sigma(n)}),$$
where we think of the symmetric group S_{n-2k} as the group of bijections of the set $\{2k+1,2k+2,...,n\}$ to itself. Thus, by averaging, there exists $\sigma \in S_{n-2k}$ for which
$$\delta(q;a_1,a_2,...,a_{2k},a_{\sigma(2k+1)},\dots,a_{\sigma(n)}) \leqslant \exp\left(-\frac{n\log(2A)}{4A\log q}\right)\frac{1}{n!},$$
which gives the first part of Theorem 1.2 on taking $C = 4A/\log(2A)$.

Similarly for the second part of the theorem, taking $k = \lceil n/2A\rceil$ we have
$$\begin{aligned}\sum_{\sigma\in S_{n-2k}} \delta(q;a_1,a_3,...,a_{2k-1},a_{\sigma(2k+1)},\dots,a_{\sigma(n)},a_{2k},...,a_4,a_2) &= \delta_{2k}^\#(q;a_1,\dots,a_n)\\ &\geqslant \exp\left(\frac{n\log(2A)}{4A\log q}\right)\frac{(n-2k)!}{n!},\end{aligned}$$
so by averaging there exists σ with
$$\delta(q;a_1,a_3,...,a_{2k-1},a_{\sigma(2k+1)},\dots,a_{\sigma(n)},a_{2k},...,a_4,a_2) \geqslant \exp\left(\frac{n\log(2A)}{4A\log q}\right)\frac{1}{n!}.$$

Finally, if $\varphi(q)^{1/50} < n \leqslant \varphi(q)$ then set $m := \lfloor \varphi(q)^{1/50} \rfloor$ and assume that q is so large that $m \geqslant 2$, hence $m \geqslant \frac{1}{2}\varphi(q)^{1/50}$. As in the previous discussion there exists an m-tuple $(a_1, \dots, a_m) \in \mathcal{A}_m(q)$ for which

$$\delta(q; a_1, \dots, a_m) \leqslant \exp\left(-\frac{\varphi(q)^{1/50}}{C \log q}\right)\frac{1}{m!},$$

and there exists an m-tuple $(b_1, \dots, b_m) \in \mathcal{A}_m(q)$ for which

$$\delta(q; b_1, \dots, b_m) \geqslant \exp\left(\frac{\varphi(q)^{1/50}}{C \log q}\right)\frac{1}{m!},$$

with $C = 8A/\log(2A)$. Then if we choose any other coprime residues $a_{m+1}, ..., a_n$ mod q, we have

$$\delta(q; a_1, \dots, a_m) = \sum_{\substack{\sigma \in S_n: \\ \sigma^{-1}(1) > \sigma^{-1}(2) > \dots > \sigma^{-1}(m)}} \delta(q; a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}).$$

There are $n!/m!$ terms in the sum, so it follows that for at least one permutation σ we must have $\delta(q; a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}) \leqslant \exp\left(-\frac{\varphi(q)^{1/50}}{C \log q}\right)\frac{1}{n!}$, as desired. The analogous lower bound with the b_i is proved exactly similarly. □

2. Key ingredients and outline of the proof of Theorem 1.3

In this section we give an outline and describe the key ideas and ingredients in the proof of Theorem 1.3. We defer the proofs of these results to later sections. Throughout the remainder of this section, we will suppose as in Theorem 1.3 that

$$1 \leqslant k \leqslant \frac{n}{A} \leqslant \varphi(q)^{1/50}, \tag{2.1}$$

where A is a sufficiently large, absolute constant. We assume that q is sufficiently large, and assume GRH and LI.

In brief, our argument has the following overall plan, which will be described in full detail in the next subsections:

(i) We express the desired logarithmic density in terms of the ordering probability of a certain vector of random variables $\mathbf{X} = (X_1, \dots, X_n)$, where each X_i has mean close to zero. This is standard given GRH and LI, and we briefly recall material from [14] and [8].

(ii) We show that the vector $\mathbf{X} = (X_1, \dots, X_n)$ can be well-approximated by a Gaussian vector $\mathbf{Z}$ with zero means and with the same covariance matrix as $\mathbf{X}$, in the region of "medium-deviations". Recalling the conclusions in Theorem 1.3, we are dealing with probabilities of order roughly $\frac{(n-2k)!}{n!} \approx n^{-2k}$, and we need to approximate the probability involving $\mathbf{X}$ closely by the corresponding ordering probability for $\mathbf{Z}$, uniformly in n and k. We accomplish this, in particular, by smoothing out the indicator function of the ordering probability in question, that is, either (1.3) or (1.4). This is a crucial ingredient, and we develop new probabilistic tools to obtain what we need.

(iii) We make a specific choice for the residue classes $a_1, \dots, a_n$. Incidentally, we utilize the *same* tuple for both parts of Theorem 1.3, and we shall explain the reasons for this shortly. With our selection, many of the correlations between the variables X_i are large in magnitude, and this is the ultimate source of the biases in our theorems. Notice this is a completely different source of bias than the influence of being a quadratic residue or non-residue, which produces Chebyshev's bias for two-way races when q is small, because the *mean values* of the corresponding X_i are slightly skewed away from zero.

As we mentioned, Harper and Lamzouri [8] proved a non-uniformity result for the auxiliary quantities $\delta_k^\flat(q; a_1, \dots, a_n)$, and then deduced their non-uniformity result for $\delta(q; a_1, \dots, a_n)$ by averaging, and we are following the same broad strategy. The advantage of looking at $\delta_k^\flat(q; a_1, \dots, a_n)$ is that in the corresponding ordering probability, namely

$$\mathbb{P}(X_1 > X_2 > ... > X_k > \max_{k+1 \leqslant j \leqslant n} X_j),$$

the maximum of the X_j (when they are normalized by their standard deviations) is close to $\sqrt{2 \log n}$ with very high probability. This means that $X_1, ..., X_k$ are all larger than $\sqrt{2 \log n}$ with very high probability, so that large correlations will have a large biasing influence on the probability. A similar effect holds for $\delta_{2k}^\sharp(a_1, \dots, a_n)$; see subsection 2.4 below for more details. The introduction of $\delta_{2k}^\sharp(a_1, \dots, a_n)$, which is another new ingredient here as compared with Harper and Lamzouri [8], is important when seeking abnormally large logarithmic densities (as in (1.6)) in addition to the abnormally small densities (as in (1.5)) provided by studying $\delta_{2k}^\flat(q; a_1, \dots, a_n)$.

(iv) Finally, we perform calculations with our smoothed function of our Gaussian vectors. Using the heuristic that $X_1, \dots, X_{2k}$ are all very large (in magnitude) with high probability, we are able to "factor out" the bias term from the corresponding density function. It remains to bound the density function of a vector of *independent* Gaussians, which is much easier to handle.

2.1. **Notation.** We adopt familiar order of magnitude notation of Bachmann–Landau, Vinogradov and Knuth. The notations $f = O(g)$, $f \ll g$ and $g \gg f$ mean that there exists a positive constant C such that $|f| \leqslant Cg$ throughout the range of f (either implicitly understood or explicitly given). The notations $f \asymp g$, $f = \Theta(g)$ mean that both $f \ll g$ and $f \gg g$ hold. We have $f(x) \sim g(x)$ as $x \to a$ if $\lim_{x\to a} f(x)/g(x) = 1$, where a may be finite, ∞ or $-\infty$.

We use $\mathbb{P}$ and $\mathbb{E}$ to denote probability and probabilistic expectation, respectively. The underlying probability spaces will be described explicitly or understood from context.

2.2. **Probabilistic expression for the logarithmic densities.** In this section, we review the connection between quantities like $\delta(q; a_1, \dots, a_n)$ and ordering probabilities for suitable random variables. See section 2 of Harper and Lamzouri [8] for a similar but more detailed review of this material.

Given distinct reduced residues $a_1, \dots, a_n$ modulo q, we define

$$E_{q;a_1,\dots,a_n}(x) := \Big(E(x; q, a_1), \dots, E(x; q, a_n)\Big),$$

where

$$E(x; q, a) := \frac{\log x}{\sqrt{x}} \left(\varphi(q)\pi(x; q, a) - \pi(x)\right),$$

and $\pi(x)$ denotes the total number of primes less than x. This normalization ensures that, if we assume GRH, then $E_{q;a_1,\dots,a_n}(x)$ varies roughly boundedly as x varies. Notice also that

$$\pi(x; q, a_1) > \cdots > \pi(x; q, a_n) \quad \Longleftrightarrow \quad E(x; q, a_1) > \cdots > E(x; q, a_n).$$

Extending the theory of Hooley [9] in the case $n = 1$, Rubinstein and Sarnak [14] showed, under GRH and LI, that the vector $E_{q;a_1,\dots,a_n}(x)$ possesses a certain limiting distribution measured on a logarithmic scale. To be precise, for any Lebesgue measurable set $T \subset \mathbb{R}^n$ whose boundary has measure zero, we have

$$\lim_{Y\to\infty} \frac{1}{\log Y} \int\limits_{\substack{x\in[2,Y] \\ E_{q;a_1,\dots,a_n}(x)\in T}} \frac{dx}{x} = \int_T d\mu_{q;a_1,\dots,a_n}, \tag{2.2}$$

where $\mu_{q;a_1,\dots,a_n}$ is a certain probability measure on $\mathbb{R}^n$ (which is absolutely continuous when $n < \varphi(q)$).

Specifically, $\mu_{q;a_1,\dots,a_n}$ is the probability measure corresponding to the $\mathbb{R}^n$-valued random vector $\big(X(q, a_1), \dots, X(q, a_n)\big)$, where

$$X(q, a) := -C_q(a) + \sum_{\substack{\chi\neq\chi_0 \\ \chi \ (\mathrm{mod}\ q)}} \mathrm{Re}\left(2\chi(a) \sum_{\gamma_\chi>0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}\right), \tag{2.3}$$

where χ runs over non-principal Dirichlet characters modulo q, γ_χ runs over the positive imaginary parts of nontrivial zeros of $L(s,\chi)$ (which are distinct by LI), $\{U(\gamma_\chi)\}_{\chi\neq\chi_0,\gamma_\chi>0}$ is a sequence of independent random variables uniformly distributed on the unit circle, and

$$C_q(a) := -1 + |\{b \pmod q : b^2 \equiv a \pmod q\}|. \tag{2.4}$$

In order to understand the probabilities and dependencies of events involving the $X(q,a_i)$, it is crucial to understand their covariances. Assuming GRH, we have (see section 2 of Harper and Lamzouri [8] for fuller details and references)

$$\mathbb{E}X(q,a_i)X(q,a_j) - \mathbb{E}X(q,a_i)\mathbb{E}X(q,a_j) = \begin{cases} \mathrm{Var}(q) & \text{if } i=j \\ B_q(a_i,a_j) & \text{if } i\neq j, \end{cases} \tag{2.5}$$

where

$$\mathrm{Var}(q) := 2 \sum_{\substack{\chi\neq\chi_0 \\ \chi \pmod q}} \sum_{\gamma_\chi>0} \frac{1}{\frac14+\gamma_\chi^2} \sim \varphi(q)\log q \qquad \text{as } q\to\infty, \tag{2.6}$$

and

$$B_q(a,b) := \sum_{\substack{\chi\neq\chi_0 \\ \chi \pmod q}} \sum_{\gamma_\chi>0} \frac{\chi\left(\frac{b}{a}\right)+\chi\left(\frac{a}{b}\right)}{\frac14+\gamma_\chi^2} \ll \varphi(q), \qquad a\neq b. \tag{2.7}$$

In light of the covariances (2.5), we define the normalized vector

$$\mathbf{X} := (X_1,\dots,X_n), \quad X_i := \frac{X(q,a_i)}{\sqrt{\mathrm{Var}(q)}}. \tag{2.8}$$

By (2.3), (2.4) and (2.6), each X_i has mean $\frac{-C_q(a_i)}{\sqrt{\mathrm{Var}(q)}}$, which is close to zero, and has variance 1.

2.3. **Approximation by a Gaussian.** Here, we will state our main approximation result, which shows that the vector of random variables $\mathbf{X}$ may be well approximated by a vector of Gaussians, in a suitable range which is applicable for the proofs of our theorems. To this end, for any n-tuple $(a_1,\dots,a_n)\in\mathcal{A}_n(q)$, let $\mathbf{Z}=\mathbf{Z}(a_1,\dots,a_n)$ be the multivariate Gaussian vector, with each component having mean zero and variance 1, and such that the correlations satisfy

$$\mathbb{E}Z_iZ_j = \frac{B_q(a_i,a_j)}{\mathrm{Var}(q)}. \tag{2.9}$$

In order to encode the conditions inherent in (1.3) and (1.4), for any set $S \subseteq \{1, 2, ..., n\} \times \{1, 2, ..., n\}$, let

$$R(S) := \{(x_1, ..., x_n) \in \mathbb{R}^n : x_i > x_j \ \forall (i, j) \in S\}. \tag{2.10}$$

We denote by $\mathbf{1}_{R(S)}(x_1, ..., x_n)$ the indicator function of the set $R(S)$. We say that the set S is *admissible* if $R(S)$ is nonempty. In particular, S does not contain any diagonal pairs (i, i). When S is admissible, we will make use of certain smooth approximations $h^+_{S,\alpha}$, $h^-_{S,\alpha}$ of the indicator function of $R(S)$ (these will depend on a parameter α, which we will take to be small). The precise definitions of $h^{\pm}_{S,\alpha}$ will be given later (see (3.1)) in section 3. We apply these with the two specific sets

$$S^\flat := \{(i, i+1) : 1 \leqslant i \leqslant 2k-1\} \cup \{(2k, j) : 2k+1 \leqslant j \leqslant n\} \tag{2.11}$$

and

$$\begin{aligned} S^\sharp := &\{(2i-1, 2i+1) : 1 \leqslant i \leqslant k-1\} \cup \{(2i+2, 2i) : 1 \leqslant i \leqslant k-1\} \\ &\cup \{(2k-1, j) : 2k \leqslant j \leqslant n\} \cup \{(i, 2k) : 2k+1 \leqslant i \leqslant n\}. \end{aligned} \tag{2.12}$$

Recalling the definitions (1.3) and (1.4), the identity (2.2) and definition (2.8), our goal is now to bound

$$\delta^\flat_{2k}(q; a_1, \ldots, a_n) = \mathbb{P}(\mathbf{X} \in R(S^\flat)), \tag{2.13}$$

and

$$\delta^\sharp_{2k}(q; a_1, \ldots, a_n) = \mathbb{P}(\mathbf{X} \in R(S^\sharp)). \tag{2.14}$$

We will take a near-optimal choice of the parameter α as

$$\alpha = \frac{1}{n^5 \log^6 q}. \tag{2.15}$$

Proposition 2.1. *Assume* (2.1) *and* (2.15). *Then for any admissible set S, we have*

$$\begin{aligned} &\left(1 + O\left(\frac{1}{\log q}\right)\right) \mathbb{E} h^-_{S,\alpha}(\mathbf{Z}) - O\left(e^{-(n \log q)^2}\right) \\ &\qquad \leqslant \mathbb{P}(\mathbf{X} \in R(S)) \leqslant \left(1 + O\left(\frac{1}{\log q}\right)\right) \mathbb{E} h^+_{S,\alpha}(\mathbf{Z}) + O\left(e^{-(n \log q)^2}\right). \end{aligned}$$

This result is a special case of Proposition 3.5. The required hypothesis $\alpha \geqslant \left(\frac{n^5 \log q}{\sqrt{\phi(q)}}\right)^{1/3}$ in Proposition 3.5 follows from (2.15) and our assumptions (2.1) on the sizes of k, n. Likewise, by (2.15) and (2.1), we have

$$ne^{-\theta(\frac{\alpha^2\phi(q)^{1/3}}{n^4\log^{2/3}q})} \leqslant e^{-(n\log q)^2}.$$

Proposition 3.5 follows from a very general multivariate approximation result, and will be proved in section 3. Here we only discuss a few salient points. Since the $X(q, a_i)$ are a sum of independent random variables, and have variance $\asymp \varphi(q)\log q$, we expect standard Berry–Esseen type ideas to produce a multivariate Gaussian approximation with an error term saving a polynomial in q, and with a polynomial dependence on the dimension n. Now comparing with Theorem 1.3, we are dealing with probabilities of size roughly $\frac{(n-2k)!}{n!} \approx n^{-2k}$ with k as large as a power of q, which is much smaller than the possible Berry–Esseen saving. To get around this, we need a multivariate Gaussian approximation with a *relative* error term rather than an absolute one, which can therefore be useful even for very improbable events. This kind of multivariate "moderate deviation" result does exist in the probabilistic literature, but we were unable to locate a result that applied to our situation without quite a lot of reworking. (See Theorem 1 of Bentkus [1] for a sample of what is available— Bentkus obtains a moderate deviation Gaussian approximation for the norm of a sum of independent, identically distributed random elements in a Hilbert space.) Thus we prove our own bespoke result (see Proposition 3.4 and Proposition 3.5, below), using a special construction of smooth test functions and an inductive "replacement" argument.

2.4. **Choosing the n-tuple and analyzing correlations.** Achieving the bounds (1.5) and (1.6) will be accomplished by making a strategic choice of the residues $a_1, ..., a_n$, for which the covariances of Z_i, Z_j have good properties. In view of (2.6) and (2.7), we have the normalized correlation bound $\frac{B_q(a,b)}{\mathrm{Var}(q)} \ll \frac{1}{\log q}$ when $a \neq b$. This upper bound can be attained by certain pairs a, b (though only when $B_q(a, b)$ is negative), and in general the behaviour of $B_q(a, b)$ depends in a complicated way on the arithmetic properties of a, b and q (see section 3 of Harper and Lamzouri [8], for example). For the application to Theorem 1.3, we will restrict to special sets of residues where the behaviour is nice.

Lemma 2.2. *Assume GRH. There is a set $\mathcal{A} = \{a_1, \ldots, a_n\}$ of distinct reduced residues modulo q, with $a_2 = -a_1$, $a_4 = -a_3, \ldots, a_{2k} = -a_{2k-1}$, $a_i \neq -a_j$ otherwise, and such that the corresponding random variables $Z_1, ..., Z_n$ satisfy the correlation bounds*

$$\mathbb{E}Z_iZ_j = \xi\, \mathbf{1}(a_i = -a_j) + O\left(\frac{n\log^3 q}{\phi(q)}\right) \qquad \forall\, i \neq j,$$

where

$$\xi := -\frac{\phi(q)\log 2}{\mathrm{Var}(q)} \sim -\frac{\log 2}{\log q}. \tag{2.16}$$

Proof. Let $b_1, ..., b_n$ be any distinct primes in the interval $(5n\log^2 q, 10n\log^2 q]$ that do not divide q. By the prime number theorem, this interval contains $\sim \frac{5n\log^2 q}{\log(10n\log^2 q)} \geqslant 4n\log q$ primes provided q is large enough, and q has at most $2\log q$ prime factors, so there are at least n primes left to choose from. Let $a_{2r-1} = b_r, a_{2r} = -b_r$ for $1 \leqslant r \leqslant k$, and $a_r = b_r$ for $2k+1 \leqslant r \leqslant n$. The claimed estimate for $\mathbb{E}Z_iZ_j$ now follows from Proposition 6.1 of Lamzouri [11] and (2.9), noting that we have $1/2 < \frac{|a|}{|a'|} < 2$ for all $a, a' \in \mathcal{A}$, so the term $\Lambda_0(\cdot)$ in that Proposition always vanishes. □

It is crucial to our argument that certain pairs of covariances are large, but we also need to ensure that any other correlations do not interfere with the resulting bias. Harper and Lamzouri [8] accomplished this using a "large sieve" kind of average estimate for correlations, but here we take a simpler approach by working with the specially chosen residues $a_1, ..., a_n$, for which all the correlations are either of size ξ or else extremely small. Since we are aiming to show extremal behaviour, rather than behaviour for all sets of residue classes, this is acceptable. As well as being simpler, this way of running the argument allows us to take k very large, compared with Harper and Lamzouri [8] who had bounded k, which is crucial to obtain very large biases. (Recall that in the deduction of Theorem 1.2 from Theorem 1.3, we take $k = \lceil n/2A \rceil$.)

We need good estimates for the density function for the vector $\mathbf{Z}$, which we denote by $f(\mathbf{x})$. For brevity, we also define the following Euclidean norms:

$$\|\mathbf{x}_{2k}\| = (x_1^2 + \cdots + x_{2k}^2)^{1/2},$$
$$\|\mathbf{x}^{n-2k}\| = (x_{2k+1}^2 + \cdots + x_n^2)^{1/2}.$$

The following result will be proved in section 4.

Proposition 2.3. *Assume* (2.1). *Let* $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. *Then*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(\xi \sum_{j=1}^{k} x_{2j-1}x_{2j} - \frac{\|\mathbf{x}_{2k}\|^2}{2}\left(1 + O\left(\frac{1}{(\log q)^2}\right)\right)\right.$$
$$\left. - \frac{\|\mathbf{x}^{n-2k}\|^2}{2}\left(1 + O\left(\frac{n^2\log^3 q}{\varphi(q)}\right)\right) + O\left(\frac{k}{(\log q)^2}\right)\right). \tag{2.17}$$

We also have the cruder estimate

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\Bigg(- \frac{\|\mathbf{x}_{2k}\|^2}{2}\left(1 + O\left(\frac{1}{\log q}\right)\right) - \frac{\|\mathbf{x}^{n-2k}\|^2}{2}\left(1 + O\left(\frac{n^2 \log^3 q}{\varphi(q)}\right)\right) + O\left(\frac{k}{(\log q)^2}\right)\Bigg). \tag{2.18}$$

2.5. **Separating out the bias term and relating to independent Gaussians.** As mentioned in the overview, when working with the logarithmic density $\delta^\flat_{2k}(q; a_1, \dots, a_n)$ via the upper bound of Proposition 2.1, the dominant contribution will come from those $\mathbf{Z}$ with $Z_1, \dots, Z_{2k} \gtrsim \sqrt{2\log n}$. This produces a "bias term" thanks to the sum in the density function (2.17), of size approximately $\sim -\frac{k\log 2}{\log q}(\sqrt{2\log n})^2 = -\frac{(2\log 2)k\log n}{\log q}$. Likewise, to bound $\delta^\sharp_{2k}(q; a_1, \dots, a_n)$ via the lower bound of Proposition 2.1, the dominant contribution will come from those $\mathbf{Z}$ with $Z_1, Z_3, \dots, Z_{2k-1} \gtrsim \sqrt{2\log n}$ and $Z_2, Z_4, \dots, Z_{2k} \lesssim -\sqrt{2\log n}$. We obtain a positive bias $\asymp -\frac{\log 2}{\log q} k\sqrt{2\log n}(-\sqrt{2\log n}) \asymp \frac{k\log n}{\log q}$ in the density (2.17) on that event. In both of these estimates, it is also important that the other correlations, corresponding to the "big Oh" term in Lemma 2.2, are uniformly very small and hence contribute negligibly.

We prove the following results in section 6.

Proposition 2.4. *Assume* (2.1) *and* (2.15). *We have*

$$\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z}) \leqslant \exp\left(-\frac{0.6k\log(n/k)}{\log q}\right)\frac{(n-2k)!}{n!}.$$

Proposition 2.5. *Assume* (2.1) *and* (2.15). *We have*

$$\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z}) \geqslant \exp\left(\frac{0.6k\log(n/k)}{\log q}\right)\frac{(n-2k)!}{n!}.$$

As remarked earlier, the main idea in the proofs of Propositions 2.4 and 2.5 is that $Z_1, ..., Z_{2k}$ are large with high probability, and this produces the desired bias in the density function (2.17). This bias effect can be "factored out" from the exponential in the density, essentially yielding the prefactors in Propositions 2.4 and 2.5, and reducing the proof to estimating $\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W})$ and $\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})$, where $\mathbf{W} = (W_1, \dots, W_n)$ is a vector of *independent* standard Gaussian random variables. By construction, the smooth functions $h^+_{S^\flat,\alpha}$ and $h^-_{S^\sharp,\alpha}$ should behave a lot like the indicator functions $\mathbf{1}_{R(S^\flat)}$ and $\mathbf{1}_{R(S^\sharp)}$ respectively. Thus, $\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W})$ and $\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})$ should both be close to $\frac{(n-2k)!}{n!}$, the probability that such a vector $\mathbf{W}$ would satisfy the orderings dictated by $S^\flat$ and $S^\sharp$ respectively. The following result shows that this is indeed the case. This corresponds to special cases of Lemmas 5.1 and 5.2, which will be proved in section 5 below.

Lemma 2.6. *Assume* (2.1) *and* (2.15). *Then*

$$\mathbb{E}h^{+}_{S^{\flat},\alpha}(\mathbf{W}) \leqslant \left(1+O\left(\frac{1}{\log^2 q}\right)\right)\frac{(n-2k)!}{n!}, \tag{2.19}$$

and

$$\mathbb{E}h^{-}_{S^{\sharp},\alpha}(\mathbf{W}), \mathbb{E}h^{+}_{S^{\sharp},\alpha}(\mathbf{W}) = \left(1+O\left(\frac{1}{\log^2 q}\right)\right)\frac{(n-2k)!}{n!}. \tag{2.20}$$

2.6. **Deducing Theorem 1.3.** With all of the above key ingredients in place, we are now ready to prove Theorem 1.3. Combining (2.13) with Propositions 2.1 and 2.4 we obtain

$$\begin{aligned}\delta^{\flat}_{2k}(q; a_1, \ldots, a_n) &\leqslant \left(1+O\left(\frac{1}{\log q}\right)\right)\mathbb{E}h^{+}_{S^{\flat},\alpha}(\mathbf{Z}) + O\left(e^{-(n\log q)^2}\right)\\ &\leqslant \left(1+O\left(\frac{1}{\log q}\right)\right)\exp\left(-\frac{0.6k\log(n/k)}{\log q}\right)\frac{(n-2k)!}{n!} + O\left(e^{-(n\log q)^2}\right).\end{aligned}$$

The second "big Oh" term here is negligibly small compared with the first term, since we have $e^{-(n\log q)^2} \leqslant n^{-n}e^{-(n\log q)^2/2} \leqslant \frac{(n-2k)!}{n!}e^{-(n\log q)^2/2}$, say. Thus we can rewrite our inequality in the form

$$\delta^{\flat}_{2k}(q; a_1, \ldots, a_n) \leqslant \exp\left(-\frac{1}{\log q}(0.6k\log(n/k) + O(1))\right)\frac{(n-2k)!}{n!}.$$

If the absolute constant A in our hypotheses (for which $n/k \geqslant A$) is fixed large enough, the right hand side will be at most $\exp\left(-\frac{k\log(n/k)}{2\log q}\right)\frac{(n-2k)!}{n!}$, which implies (1.5).

The proof of (1.6) follows along the same lines by using (2.14), along with Propositions 2.1 and 2.5.

3. Gaussian approximation

In this section, we prove the results that will allow us to replace the actual random variables $X(q, a_i)$ arising from prime number races by Gaussian random variables with the same covariance structure. As explained in section 2, this will be important because jointly Gaussian random variables have an explicit probability density, depending only on the means and covariances, which we can work with to analyse the probabilities of events.

3.1. **Smooth test functions.** We begin by constructing smooth weight functions that closely approximate indicators of the "ordering" events that we are interested in. Smooth weights will allow us to use Taylor expansion as a tool when we come to our Gaussian approximation.

Lemma 3.1. *There exists an increasing function $\theta : \mathbb{R} \to (0,1)$ that is three times continuously differentiable, and satisfies*

$$\theta(-x) \ll e^{-x}, \qquad 1-\theta(x) \ll e^{-x} \qquad \forall x \geqslant 0,$$

as well as

$$|\frac{d^m}{dx^m}\theta(x)| \ll \theta(x) \quad \forall 1 \leqslant m \leqslant 3,\ \forall x \in \mathbb{R}.$$

Proof of Lemma 3.1. We take $\theta(x)$ to be the probability distribution function corresponding to a density proportional to $e^{-|x|}$, since repeatedly differentiating the exponential continues to yield an exponential. Let $f(x) = e^{-|x|}$ for $|x| \geqslant 1$, and when $|x| < 1$, let $f(x)$ be an appropriate quadratic polynomial which ensures that $f \in C^2(\mathbb{R})$; the polynomial $\frac{7-4x^2+x^4}{4e}$ satisfies these requirements. If $m_0 := \int_{-\infty}^{\infty} f(t)dt$, then clearly

$$\theta(x) := \frac{1}{m_0}\int_{-\infty}^{x} f(t)dt$$

satisfies all the required properties of the lemma. In particular, for $x \leqslant -1$, we have $|\theta^{(j)}(x)| = \theta(x) = \frac{1}{m_0}e^x$ for $j = 1,2,3$ and for $x \geqslant -1$, $\theta(x) \geqslant 1/(em_0) \gg |\theta^{(j)}(x)|$ for $j = 1,2,3$. □

By taking a product of (shifted and dilated) copies of the function $\theta(x)$, we can obtain multi-dimensional smooth test functions, which is what we shall actually need. Recall that a set $S \subseteq \{1,2,...,n\} \times \{1,2,...,n\}$ is admissible if

$$R(S) = \{(x_1,...,x_n) \in \mathbb{R}^n : x_i > x_j\ \forall (i,j) \in S\}$$

is nonempty. For an admissible set S and $\alpha > 0$ we define the two functions $h^+_{S,\alpha}$ and $h^-_{S,\alpha}$ by

$$h^{\pm}_{S,\alpha}(x_1,...,x_n) := \prod_{(i,j)\in S} g(x_i - x_j \pm \sqrt{\alpha}), \qquad \text{where } g(x) := \theta(x/\alpha). \tag{3.1}$$

We next show that these functions are good approximations of $\mathbf{1}_{R(S)}(x_1,...,x_n)$.

Proposition 3.2. *Let S be an admissible set. Each of $h^-_{S,\alpha}$ and $h^+_{S,\alpha}$ is a three times continuously differentiable function from $\mathbb{R}^n$ to $[0,\infty)$, satisfying*

$$h^{\pm}_{S,\alpha}(\boldsymbol{x}) \leqslant 1, \quad \sup_{1\leqslant i\leqslant n}\left|\frac{\partial}{\partial x_i}h^{\pm}_{S,\alpha}(\boldsymbol{x})\right| \ll \frac{n}{\alpha}h^{\pm}_{S,\alpha}(\boldsymbol{x}), \quad \sup_{1\leqslant i,j,k\leqslant n}\left|\frac{\partial^3}{\partial x_i\partial x_j\partial x_k}h^{\pm}_{S,\alpha}(\boldsymbol{x})\right| \ll \left(\frac{n}{\alpha}\right)^3 h^{\pm}_{S,\alpha}(\boldsymbol{x}).$$

Furthermore, we have

$$h^-_{S,\alpha}(x_1,...,x_n) - O(e^{-1/\sqrt{\alpha}}) \leqslant \boldsymbol{1}_{R(S)}(x_1,...,x_n) \leqslant h^+_{S,\alpha}(x_1,...,x_n) + O(n^2 e^{-1/\sqrt{\alpha}}).$$

An important feature of this result is the fact that the derivatives of $h^{\pm}_{S,\alpha}$ are always controlled by $h^{\pm}_{S,\alpha}$ itself, even at points where $h^{\pm}_{S,\alpha}$ is very small. We will exploit this later in our Gaussian approximation.

Proof of Proposition 3.2. The partial derivative bounds in Proposition 3.2 follow exactly as in Lemma 4.3 of Harper and Lamzouri [8], for example, by repeated application of the product rule together with the fact that $\frac{d^m}{dx^m}g(x) = (1/\alpha)^m \frac{d^m}{dx^m}\theta(x/\alpha) \ll (1/\alpha)^m\theta(x/\alpha) = (1/\alpha)^m g(x)$ for $1 \leqslant m \leqslant 3$.

For the lower bound on $\mathbf{1}_{R(S)}$, it will suffice to show that whenever $(x_1,...,x_n) \notin R(S)$ we have a uniform upper bound

$$h^-_{S,\alpha}(x_1,...,x_n) \ll e^{-1/\sqrt{\alpha}}.$$

Indeed, if $(x_1,...,x_n) \notin R(S)$ then for at least one pair $(i,j) \in S$, we have $x_i - x_j - \sqrt{\alpha} < -\sqrt{\alpha}$, and therefore

$$h^-_{S,\alpha}(x_1,...,x_n) \leqslant g(-\sqrt{\alpha}) = \theta(-1/\sqrt{\alpha}) \ll e^{-1/\sqrt{\alpha}}.$$

Similarly, for the upper bound it will suffice to show that whenever $(x_1,...,x_n) \in R(S)$ we have

$$h^+_{S,\alpha}(x_1,...,x_n) = 1 - O(n^2 e^{-1/\sqrt{\alpha}}).$$

We may assume that $n^2 e^{-1/\sqrt{\alpha}} \leqslant 1$, otherwise the assertion is trivial. And indeed, if $(x_1,...,x_n) \in R(S)$ then we have $x_i - x_j + \sqrt{\alpha} \geqslant \sqrt{\alpha}$ for each pair $(i,j) \in S$, and so

$$h^+_{S,\alpha}(x_1,...,x_n) \geqslant g(\sqrt{\alpha})^{\#S} \geqslant \theta(1/\sqrt{\alpha})^{n^2},$$

which implies the result since we have $\theta(1/\sqrt{\alpha}) = 1 - O(e^{-1/\sqrt{\alpha}})$.

□

3.2. **A Lindeberg type argument.** In this subsection, we shall establish a "moderate deviation" type of Gaussian approximation result relative to smooth test functions h. By "moderate deviation", we mean that we want the theorem to remain useful even in the tails of the Gaussian, where an approximation with a simple absolute error term would not be useful. Instead, we are seeking a result with a relative error (together with an absolute error term that is extremely small).

We shall obtain our Gaussian approximation with a version of the Lindeberg replacement strategy, which was originally used to prove the classical central limit theorem for sums of independent random variables. The idea is to Taylor expand the test function h to third order, and replace the independent summands $\mathbf{X}^{(j)}$ by corresponding Gaussian random variables $\mathbf{Z}^{(j)}$ one at a time. The slightly non-standard assumptions we shall make on h, that its partial derivatives are controlled by h itself, will allow us to make the biggest error term a relative rather than absolute one. (So far as we are aware, the use of such non-standard h is novel in this context. The use of a Lindeberg type method is certainly not novel, for example this is how Bentkus [1] proceeds.)

Let $C_1, C_3 \geqslant 1$, and let $h : \mathbb{R}^n \to [0, \infty)$ be a three times continuously differentiable function, such that for all $\mathbf{x} \in \mathbb{R}^n$ we have

$$h(\mathbf{x}) \leqslant 1, \quad \sup_{1 \leqslant i \leqslant n} \left| \frac{\partial}{\partial x_i} h(\mathbf{x}) \right| \leqslant C_1 h(\mathbf{x}), \quad \sup_{1 \leqslant i,j,k \leqslant n} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h(\mathbf{x}) \right| \leqslant C_3 h(\mathbf{x}).$$

Lemma 3.3. *Let $0 < \epsilon \leqslant \min\{1/(2C_1), 1/(2C_3^{1/3})\}$. Then uniformly for any fixed $\mathbf{Y} \in \mathbb{R}^n$, any real random vector $\mathbf{X} = (X_1, ..., X_n)$ whose components have mean zero, we have the following. If $\mathbf{Z} = (Z_1, ..., Z_n)$ is a multivariate normal random vector whose components have mean zero and the same covariances $r_{i,j} := \mathbb{E} X_i X_j$ as $\mathbf{X}$, then*

$$\begin{aligned} \mathbb{E}h(\mathbf{Y}+\mathbf{X}) &= (1+\gamma(\epsilon))\mathbb{E}h(\mathbf{Y}+\mathbf{Z}) + \\ &\quad + O\Bigg(C_3 \Bigg(\sum_{i=1}^n \mathbb{E}\Big(\mathbf{1}_{|X_i| > \epsilon/n} (\epsilon^3 + n^2 |X_i|^3) \Big) + \sum_{i=1}^n (\epsilon^3 + n^2 r_{i,i}^{3/2}) e^{-\epsilon^2/(2n^2 r_{i,i})} \Bigg) \Bigg), \end{aligned}$$

where $\gamma(\epsilon) = \gamma(\epsilon, h, \mathbf{Y}, \mathbf{X})$ satisfies $|\gamma(\epsilon)| \leqslant 6C_3\epsilon^3$.

Proof of Lemma 3.3. By the multivariate form of Taylor's theorem, we have

$$h(\mathbf{Y}+\mathbf{X}) = h(\mathbf{Y}) + \sum_{i=1}^n X_i \frac{\partial}{\partial x_i} h(\mathbf{Y}) + \sum_{i,j=1}^{n}{}^{*} X_i X_j \frac{\partial^2}{\partial x_i \partial x_j} h(\mathbf{Y}) + R(h, \mathbf{Y}, \mathbf{X}),$$

where the $*$ on the sum indicates that the terms $i = j$ should be counted with weight $1/2$, and where the error term satisfies

$$\begin{aligned} |R(h, \mathbf{Y}, \mathbf{X})| &\leqslant \sup_{\theta \in [0,1]} \sup_{1 \leqslant i,j,k \leqslant n} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h(\mathbf{Y} + \theta \mathbf{X}) \right| \Big(\sum_{i=1}^n |X_i| \Big)^3 \\ &\leqslant C_3 \sup_{\theta \in [0,1]} h(\mathbf{Y}+\theta\mathbf{X}) \Big(\sum_{i=1}^n |X_i| \Big)^3. \end{aligned}$$

One has the same expansion for $h(\mathbf{Y}+\mathbf{Z})$. Taking expectations and then taking the difference $\mathbb{E}h(\mathbf{Y}+\mathbf{X}) - \mathbb{E}h(\mathbf{Y}+\mathbf{Z})$, all of the main terms cancel because we assume $\mathbf{Z}$ and $\mathbf{X}$ have the same means and covariances, so we get

$$\begin{aligned}|\mathbb{E}h(\mathbf{Y}+\mathbf{X})-\mathbb{E}h(\mathbf{Y}+\mathbf{Z})| &\leqslant \mathbb{E}|R(h,\mathbf{Y},\mathbf{X})|+\mathbb{E}|R(h,\mathbf{Y},\mathbf{Z})| \\ \leqslant\ C_3\Bigg(\mathbb{E}\Bigg\{\sup_{\theta\in[0,1]} h(\mathbf{Y}+\theta\mathbf{X})\Big(\sum_{i=1}^n |X_i|\Big)^3\Bigg\}&+\mathbb{E}\Bigg\{\sup_{\theta\in[0,1]} h(\mathbf{Y}+\theta\mathbf{Z})\Big(\sum_{i=1}^n |Z_i|\Big)^3\Bigg\}\Bigg).\end{aligned}$$

Now using Taylor's theorem again, for any $\theta\in[0,1]$ we have

$$\begin{aligned}|h(\mathbf{Y}+\theta\mathbf{X})-h(\mathbf{Y}+\mathbf{X})| &\leqslant \sup_{\phi\in[0,1]}\sup_{1\leqslant i\leqslant n}\left|\frac{\partial}{\partial x_i}h(\mathbf{Y}+\phi\mathbf{X})\right|\sum_{i=1}^n |X_i| \\ &\leqslant C_1\sup_{\phi\in[0,1]} h(\mathbf{Y}+\phi\mathbf{X})\sum_{i=1}^n |X_i|.\end{aligned}$$

In particular, if we happen to have $\sum_{i=1}^n |X_i|\leqslant \epsilon\leqslant 1/(2C_1)$ then it follows that

$$\sup_{\theta\in[0,1]} h(\mathbf{Y}+\theta\mathbf{X})\leqslant 2h(\mathbf{Y}+\mathbf{X}).$$

So we always have the bound

$$\mathbb{E}\sup_{\theta\in[0,1]} h(\mathbf{Y}+\theta\mathbf{X})\Big(\sum_{i=1}^n |X_i|\Big)^3\leqslant 2\epsilon^3\mathbb{E}h(\mathbf{Y}+\mathbf{X})+\mathbb{E}\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon}\Big(\sum_{i=1}^n |X_i|\Big)^3.$$

The same argument applies to $\sup_{\theta\in[0,1]} h(\mathbf{Y}+\theta\mathbf{Z})$.

Putting everything together, we get

$$\begin{aligned}\big|\mathbb{E}h(\mathbf{Y}+\mathbf{X})-\mathbb{E}h(\mathbf{Y}+\mathbf{Z})\big| &\leqslant 2C_3\epsilon^3\big(\mathbb{E}h(\mathbf{Y}+\mathbf{X})+\mathbb{E}h(\mathbf{Y}+\mathbf{Z})\big)+ \\ &+C_3\left(\mathbb{E}\left(\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon}\Big(\sum_{i=1}^n |X_i|\Big)^3\right)+\mathbb{E}\left(\mathbf{1}_{\sum_{i=1}^n |Z_i|>\epsilon}\Big(\sum_{i=1}^n |Z_i|\Big)^3\right)\right).\end{aligned}$$

Furthermore, our assumptions on ϵ imply that $2C_3\epsilon^3\leqslant 1/4$, and so

$$\frac{1+2C_3\epsilon^3}{1-2C_3\epsilon^3}=1+\frac{4C_3\epsilon^3}{1-2C_3\epsilon^3}\leqslant 1+6C_3\epsilon^3,$$

and similarly $\frac{1-2C_3\epsilon^3}{1+2C_3\epsilon^3}\geqslant 1-4C_3\epsilon^3$. So rearranging our above bound, we find that

$$\begin{aligned}&\mathbb{E}h(\mathbf{Y}+\mathbf{X})=(1+\gamma(\epsilon))\mathbb{E}h(\mathbf{Y}+\mathbf{Z})+ \\ &\qquad O\Bigg(C_3\Bigg(\mathbb{E}\Big(\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon}(\sum_{i=1}^n |X_i|)^3\Big)+\mathbb{E}\Big(\mathbf{1}_{\sum_{i=1}^n |Z_i|>\epsilon}(\sum_{i=1}^n |Z_i|)^3\Big)\Bigg)\Bigg),\end{aligned}$$

where $|\gamma(\epsilon)|\leqslant 6C_3\epsilon^3$.

Next we want to work with the "big Oh" term a little, to replace terms depending on $\sum_{i=1}^n |X_i|$ and $\sum_{i=1}^n |Z_i|$ by terms that only require information about individual components X_i and Z_i, as in the statement of the lemma. First, Hölder's inequality implies that $(\sum_{i=1}^n |X_i|)^3 \leqslant n^2 \sum_{i=1}^n |X_i|^3$. Furthermore, by splitting into cases according as each $|X_i| > \epsilon/n$ or not, we get

$$\begin{aligned} n^2\mathbb{E}\Big(\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon} \sum_{i=1}^n |X_i|^3\Big) &\leqslant n^2\mathbb{E}\left(\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon} \sum_{i=1}^n \left((\epsilon/n)^3 + \mathbf{1}_{|X_i|>\epsilon/n}|X_i|^3\right)\right) \\ &\leqslant n^2 \sum_{i=1}^n \mathbb{E}\Big(\mathbf{1}_{|X_i|>\epsilon/n}|X_i|^3\Big) + \epsilon^3\mathbb{E}\big(\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon}\big). \end{aligned}$$

And we can bound this further using the inequality

$$\mathbb{E}\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon} \leqslant \mathbb{E}\mathbf{1}_{\max_i |X_i|>\epsilon/n} \leqslant \sum_{i=1}^n \mathbb{E}\mathbf{1}_{|X_i|>\epsilon/n},$$

so overall we have

$$\mathbb{E}\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon}(\sum_{i=1}^n |X_i|)^3 \leqslant n^2\mathbb{E}\mathbf{1}_{\sum_{i=1}^n |X_i|>\epsilon} \sum_{i=1}^n |X_i|^3 \leqslant \sum_{i=1}^n \mathbb{E}\mathbf{1}_{|X_i|>\epsilon/n}(\epsilon^3 + n^2|X_i|^3).$$

We have the analogous bound for the term involving the Z_i.

Finally, in the case of the Z_i, since we know that $Z_i \sim N(0, r_{i,i})$ we can further say that $\mathbb{E}(\mathbf{1}_{|Z_i|>\epsilon/n}) = \mathbb{P}(|Z_i| > \epsilon/n) \ll e^{-\epsilon^2/(2n^2 r_{i,i})}$, and that

$$\mathbb{E}\left(\mathbf{1}_{|Z_i|>\epsilon/n}|Z_i|^3\right) \ll r_{i,i}^{3/2}\left(1 + (\epsilon/n\sqrt{r_{i,i}})^3\right) e^{-\epsilon^2/(2n^2 r_{i,i})} \ll \left(r_{i,i}^{3/2} + \frac{\epsilon^3}{n^3}\right) e^{-\epsilon^2/(2n^2 r_{i,i})}.$$

Inserting these estimates in the "big Oh" term completes the proof. □

Applying Lemma 3.3 inductively, we shall prove our Gaussian approximation result for sums of m independent random vectors.

Proposition 3.4. *For $1 \leqslant j \leqslant m$, let $\mathbf{X}^{(j)} = (X_1^{(j)}, ..., X_n^{(j)})$ be independent $\mathbb{R}^n$-valued random vectors whose components have mean zero, and let $\mathbf{Z}^{(j)}$, $1 \leqslant j \leqslant m$, be independent multivariate normal random vectors whose components have mean zero and the same covariances $r(j)_{i,k} = \mathbb{E}X_i^{(j)}X_k^{(j)}$ as $\mathbf{X}^{(j)}$.*

Let $0 < \epsilon \leqslant \min\{1/(2C_1), 1/(3C_3^{1/3}m^{1/3})\}$ be a small parameter.

Then we have

$$\begin{aligned}\mathbb{E}h\Big(\sum_{j=1}^{m}\mathbf{X}^{(j)}\Big) &= e^{\Delta(\epsilon)}\mathbb{E}h\Big(\sum_{j=1}^{m}\mathbf{Z}^{(j)}\Big)+O\left(C_3\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{E}\Big(\mathbf{1}_{|X_i^{(j)}|>\epsilon/n}\big(\epsilon^3+n^2|X_i^{(j)}|^3\big)\Big)\right)\\ &\quad +O\left(C_3\sum_{j=1}^{m}\sum_{i=1}^{n}(\epsilon^3+n^2r(j)_{i,i}^{3/2})e^{-\epsilon^2/(2n^2r(j)_{i,i})}\right),\end{aligned}$$

where $\Delta(\epsilon)=\Delta_m(\epsilon,h,\{\mathbf{X}^{(j)}\})$ *satisfies* $|\Delta(\epsilon)|\leqslant 12mC_3\epsilon^3$.

To get an idea of the potential usefulness of this result, the reader might consider the case where all components $X_i^{(j)}$ have variance $\asymp 1/m$ (and so all components of the sum $\sum_{j=1}^{m}\mathbf{X}^{(j)}$ have variance $\asymp 1$), and C_1, C_3, n are fairly small compared with m. Because $\Delta(\epsilon)$ decays cubically with ϵ, any choice of ϵ that is rather smaller than $1/(mC_3)^{1/3}$ will make the relative error term $e^{\Delta(\epsilon)}$ close to 1. Meanwhile, if the random components $X_i^{(j)}$ are reasonably concentrated on the order of their standard deviations, we expect any choice of ϵ rather larger than $n/\sqrt{m}$ to yield substantial savings in the "big Oh" terms. So we have room to make a choice of ϵ that simultaneously controls all these terms.

Proof of Proposition 3.4. We proceed by induction on m. Our inductive hypothesis will be the estimate stated in the proposition, with an additional multiplier $e^{6C_3\epsilon^3(m-1)}$ in the "big Oh" terms. If we can prove this inductively we will be done, since our conditions on m imply this factor is $\leqslant e^{6/27}\ll 1$.

When $m=1$, Proposition 3.4 is a direct consequence of Lemma 3.3, with $\mathbf{Y}$ taken as the 0 vector.

For the inductive step, if we first take $\mathbf{Y}=\sum_{j=1}^{m-1}\mathbf{X}^{(j)}$ and condition on the value of $\mathbf{Y}$ (which is independent of $\mathbf{X}^{(m)}$ and $\mathbf{Z}^{(m)}$), then Lemma 3.3 implies that

$$\begin{aligned}\mathbb{E}h(\mathbf{Y}+\mathbf{X}^{(m)}) &= (1+\gamma(\epsilon))\mathbb{E}h(\mathbf{Y}+\mathbf{Z}^{(m)})+O\left(C_3\sum_{i=1}^{n}\mathbb{E}\Big(\mathbf{1}_{|X_i^{(m)}|>\epsilon/n}(\epsilon^3+n^2|X_i^{(m)}|^3)\Big)\right)\\ &\quad +O\left(C_3\sum_{i=1}^{n}(\epsilon^3+n^2r(m)_{i,i}^{3/2})e^{-\epsilon^2/(2n^2r(m)_{i,i})}\right).\end{aligned}$$

Now to understand the expectation $\mathbb{E}h(\mathbf{Y}+\mathbf{Z}^{(m)})=\mathbb{E}h(\sum_{j=1}^{m-1}\mathbf{X}^{(j)}+\mathbf{Z}^{(m)})$, if we condition on the value of $\mathbf{Z}^{(m)}$ and apply the inductive hypothesis with $h(\cdot)$ replaced by $h(\cdot+\mathbf{Z}^{(m)})$

(which obeys all the same partial derivative bounds), we get that $\mathbb{E}h(\sum_{j=1}^{m-1}\mathbf{X}^{(j)}+\mathbf{Z}^{(m)})$ is

$$e^{\Delta_{m-1}(\epsilon)}\mathbb{E}h(\sum_{j=1}^{m-1}\mathbf{Z}^{(j)}+\mathbf{Z}^{(m)})+O\left(e^{6C_3\epsilon^3(m-2)}C_3\sum_{j=1}^{m-1}\sum_{i=1}^{n}\mathbb{E}\mathbf{1}_{|X_i^{(j)}|>\epsilon/n}(\epsilon^3+n^2|X_i^{(j)}|^3)\right)$$
$$+O\left(e^{6C_3\epsilon^3(m-2)}C_3\sum_{j=1}^{m-1}\sum_{i=1}^{n}(\epsilon^3+n^2r(j)_{i,i}^{3/2})e^{-\epsilon^2/(2n^2r(j)_{i,i})}\right).$$

The above is then multiplied by $(1+\gamma(\epsilon))$ in the expression for $\mathbb{E}h(\mathbf{Y}+\mathbf{X}^{(m)})$. Using the fact that

$$e^{-12C_3\epsilon^3}\leqslant e^{-2|\gamma(\epsilon)|}\leqslant(1+\gamma(\epsilon))\leqslant e^{|\gamma(\epsilon)|}\leqslant e^{6C_3\epsilon^3},$$

we complete the induction. (Notice that when we estimate the “big Oh” terms, we only need to multiply by the upper bound $e^{6C_3\epsilon^3}$ for $(1+\gamma(\epsilon))$.) □

3.3. **Application to prime number races.** In this subsection, we specialize the discussion in the preceding propositions to the case of prime number races. Recall the random variables

$$X(q,a)=-C_q(a)+\sum_{\substack{\chi\neq\chi_0\\ \chi \pmod q}}\operatorname{Re}\left(2\chi(a)\sum_{\gamma_\chi>0}\frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4}+\gamma_\chi^2}}\right)$$

from section 2.2, and recall the sets $R(S)=\{(x_1,...,x_n)\in\mathbb{R}^n : x_i>x_j\ \forall(i,j)\in S\}\subseteq\mathbb{R}^n$ and the corresponding smooth test functions $h_{S,\alpha}^-, h_{S,\alpha}^+$ from section 3.1.

Proposition 3.5. *Let q be large, and suppose that $2\leqslant n\leqslant\varphi(q)^{1/12}$. Define $\mathbf{X}$ by (2.8), where $(a_1,...,a_n)\in\mathcal{A}_n(q)$, and let $\mathbf{Z}=(Z_i)_{1\leqslant i\leqslant n}$ denote a multivariate normal random vector whose components have mean zero, variance one, and correlations given by (2.9). Then for any small parameter $(\frac{n^5\log q}{\sqrt{\varphi(q)}})^{1/3}\leqslant\alpha$ and any admissible set S, we have*

$$\left(1+O\left(\frac{1}{\log q}\right)\right)\mathbb{E}h_{S,\alpha}^-(\mathbf{Z})-O\left(e^{-2/\sqrt{\alpha}}+ne^{-\Theta\left(\frac{\alpha^2\varphi(q)^{1/3}}{n^4\log^{2/3}q}\right)}\right)$$
$$\leqslant\mathbb{P}(\mathbf{X}\in R(S))\leqslant\left(1+O\left(\frac{1}{\log q}\right)\right)\mathbb{E}h_{S,\alpha}^+(\mathbf{Z})+O\left(n^2e^{-2/\sqrt{\alpha}}+ne^{-\Theta\left(\frac{\alpha^2\varphi(q)^{1/3}}{n^4\log^{2/3}q}\right)}\right).$$

We remark that the condition $n\leqslant\varphi(q)^{1/12}$ is stronger than necessary, but is convenient at one point in the proof, and without it the proposition is trivial because the “big Oh” term will not be less than 1.

Proof of Proposition 3.5. In the first place, we apply Proposition 3.2 with α replaced by $\alpha/4$, which gives

$$\mathbb{E}h^-_{S,\alpha/4}(\mathbf{X}) - O(e^{-2/\sqrt{\alpha}}) \leqslant \mathbb{P}(\mathbf{X} \in R(S)) \leqslant \mathbb{E}h^+_{S,\alpha/4}(\mathbf{X}) + O(n^2 e^{-2/\sqrt{\alpha}}).$$

Now the shifts $D_i := \frac{C_q(a_i)}{\sqrt{\mathrm{Var}(q)}}$ in the components $X_i := \frac{X(q,a_i)}{\sqrt{\mathrm{Var}(q)}}$ (see (2.3)) are a little awkward, since they cause the components to have non-zero mean. However, recalling our expressions for $C_q(a)$ (2.4) and $\mathrm{Var}(q)$ (2.6), together with the lower bound $\alpha \geqslant (\frac{n^5 \log q}{\sqrt{\varphi(q)}})^{1/3} \geqslant \frac{1}{\varphi(q)^{1/6}}$, we find that

$$|D_i| \leqslant \frac{\sum_{m|q} 1}{\sqrt{\mathrm{Var}(q)}} \ll \frac{1}{\varphi(q)^{0.49}} \leqslant \frac{\sqrt{\alpha}}{4}.$$

It follows that for any pair (i,j) we have

$$X_i - X_j + \sqrt{\alpha/4} \leqslant \widetilde{X}_i - \widetilde{X}_j + \sqrt{\alpha}$$

and

$$X_i - X_j - \sqrt{\alpha/4} \geqslant \widetilde{X}_i - \widetilde{X}_j - \sqrt{\alpha},$$

where

$$\widetilde{X}_i = X_i + D_i \qquad (1 \leqslant i \leqslant n).$$

So recalling the definitions (3.1) of $h^+_{S,\alpha/4}, h^-_{S,\alpha/4}$, we may remove the shifts D_i and still have the same upper and lower bounds for $\mathbb{P}(\mathbf{X} \in R(S))$, at the cost of replacing $h^{\pm}_{S,\alpha/4}$ by $h^{\pm}_{S,\alpha}$. That is, we have

$$\mathbb{E}h^-_{S,\alpha}(\widetilde{\mathbf{X}}) - O(e^{-2/\sqrt{\alpha}}) \leqslant \mathbb{P}(\mathbf{X} \in R(S)) \leqslant \mathbb{E}h^+_{S,\alpha}(\widetilde{\mathbf{X}}) + O(n^2 e^{-2/\sqrt{\alpha}}).$$

Now we want to show that $\mathbb{E}h^+_{S,\alpha}(\widetilde{\mathbf{X}})$ may be replaced by $\mathbb{E}h^+_{S,\alpha}(\mathbf{Z})$, and $\mathbb{E}h^-_{S,\alpha}(\widetilde{\mathbf{X}})$ by $\mathbb{E}h^-_{S,\alpha}(\mathbf{Z})$, up to acceptable error terms. We apply Proposition 3.4 with the sum over $1 \leqslant j \leqslant m$ replaced by a sum over characters $\chi \neq \chi_0 \bmod q$ (so $m = \varphi(q) - 1$), and with $\widetilde{X}_j$ replaced by

$$\widetilde{X}^{(\chi)} := \left(\frac{1}{\sqrt{\mathrm{Var}(q)}} \mathrm{Re} \left(2\chi(a_i) \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} \right) \right)_{1 \leqslant i \leqslant n}.$$

Notice that these are indeed independent $\mathbb{R}^n$-valued random vectors whose components have mean zero, since the underlying random variables $U(\gamma_\chi)$ are independent and have mean zero. For the test functions $h^{\pm}_{S,\alpha}$, we may take $C_1 \asymp n/\alpha$ and $C_3 \asymp (n/\alpha)^3$, so in Proposition 3.4 we are permitted to make any choice of $0 < \epsilon \leqslant \alpha/(n\varphi(q)^{1/3})$.

Turning to the error terms in Proposition 3.4, for each $1 \leqslant i \leqslant n$ and each $\chi \neq \chi_0$ we have

$$r(\chi)_{i,i} = \frac{1}{\mathrm{Var}(q)} \sum_{\gamma_\chi > 0} \frac{\mathbb{E}(\mathrm{Re} 2\chi(a_i) U(\gamma_\chi))^2}{\frac{1}{4} + \gamma_\chi^2} = \frac{2}{\mathrm{Var}(q)} \sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \ll \frac{1}{\varphi(q)},$$

where the final inequality uses the standard estimate $\sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \ll \log q$ (see Corollary 10.18 of Montgomery and Vaughan [13], for example). We remark that at this step we only need an upper bound for $r(\chi)_{i,i}$ because, as this component variance gets smaller, the corresponding upper bounds for tail probabilities will only get stronger. Furthermore, an exponential moment calculation with $\widetilde{X}_i^{(\chi)}$ (as in the proof of Lemma 2.3 of Lamzouri [10]) implies that

$$\mathbb{P}(|\widetilde{X}_i^{(\chi)}| > r) \ll e^{-\Theta(r^2 \varphi(q))}$$

for each $1 \leqslant i \leqslant n$, each $\chi \neq \chi_0$ and any $r \geqslant 0$. This simply says that, as we might expect, the components $\widetilde{X}_i^{(\chi)}$ have Gaussian-type tails. A consequence of this bound is

$$\mathbb{E}\mathbf{1}_{|\widetilde{X}_i^\chi| > r} |\widetilde{X}_i^\chi|^3 \ll (r^3 + \frac{1}{\varphi(q)^{3/2}}) e^{-\Theta(r^2 \varphi(q))}.$$

Inserting all this information in Proposition 3.4, we get for any $0 < \epsilon \leqslant \alpha/(n\varphi(q)^{1/3})$ that

$$\mathbb{E} h_{S,\alpha}^{\pm}(\widetilde{\mathbf{X}}) = e^{\Delta(\epsilon)} \mathbb{E} h_{S,\alpha}^{\pm}(\mathbf{Z}) + O\left((\frac{n}{\alpha})^3 \varphi(q) n \left(\epsilon^3 + \frac{n^2}{\varphi(q)^{3/2}} \right) e^{-\Theta(\epsilon^2 \varphi(q)/n^2)} \right),$$

where $|\Delta(\epsilon)| \ll \varphi(q)(n\epsilon/\alpha)^3$.

Finally, if we take $\epsilon = \frac{\alpha}{n\varphi(q)^{1/3} \log^{1/3} q}$ then, in view of the condition $\alpha \geqslant (\frac{n^5 \log q}{\sqrt{\varphi(q)}})^{1/3}$, we have $\epsilon^3 \geqslant \frac{n^2}{\varphi(q)^{3/2}}$. Then a quick calculation verifies that our "big Oh" term is of the form claimed in the proposition. □

4. Bounding the density of jointly Gaussian variables

We shall consider the tuple of residues $(a_1, \dots, a_n)$ satisfying the conclusion of Lemma 2.2. Our goal is to calculate explicit expressions for the probability density function of the vector $\mathbf{Z} = (Z_i)_{1 \leqslant i \leqslant n}$, where the Z_i are the jointly Gaussian random variables from Proposition 3.5 that correspond to the residues a_i.

The covariance matrix of $\mathbf{Z}$ is $C = C(\mathbf{Z}) = A + E$, where

$$A = \begin{bmatrix} 1 & \xi & & & & & & & & & & \\ \xi & 1 & & & & & & & & & & \\ & & 1 & \xi & & & & & & & & \\ & & \xi & 1 & & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & & 1 & \xi & & & & & \\ & & & & & \xi & 1 & & & & & \\ & & & & & & & 1 & & & & \\ & & & & & & & & 1 & & & \\ & & & & & & & & & \ddots & & \\ & & & & & & & & & & 1 & \\ & & & & & & & & & & & 1 \end{bmatrix},$$

ξ is defined in (2.16), and all of the entries of E are *uniformly small*, in fact bounded by ε, where

$$\varepsilon \ll \frac{n \log^3 q}{\varphi(q)} \ll \frac{\sqrt{\log q}}{\sqrt{\varphi(q)}}. \tag{4.1}$$

Let $f(x_1, \cdots, x_n)$ be the joint density function of $Z_1, \ldots, Z_n$. Since the Z_i are jointly Gaussian, we have

$$f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \exp\Big\{ -\frac{1}{2}\mathbf{x}^T C^{-1} \mathbf{x} \Big\}. \tag{4.2}$$

To establish Proposition 2.3, we first prove the following lemma.

Lemma 4.1. *Suppose q is sufficiently large that $|\xi| \leqslant 1/2$ and $\varepsilon \leqslant \frac{1}{4n}$. Then*

- $\det C = (\det A)(1 + O(\varepsilon n))$;
- *C is invertible, and $C^{-1} = A^{-1} + F$, where the entries of F are bounded in absolute value by 8ε.*

Proof. Write $C = A + E = A(I - E')$, where $E' = -A^{-1}E$. As $|\xi| \leqslant 1/2$, and the inverse of $\left(\begin{smallmatrix} 1 & \xi \\ \xi & 1 \end{smallmatrix}\right)$ is $(1-\xi^2)^{-1}\left(\begin{smallmatrix} 1 & -\xi \\ -\xi & 1 \end{smallmatrix}\right)$, we easily see that the entries of E' are bounded in absolute value by 2ε. Then $\det C = (\det A)\det(I - E')$ and, writing the determinant as a sum over permutations $\sigma \in S_n$, with t the number of fixed points of σ,

$$\begin{aligned} \det(I - E') &= (1 + O(\varepsilon))^n + O\Bigg(\sum_{\mathrm{id} \neq \sigma \in S_n} (2\varepsilon)^{n-t}(1+2\varepsilon)^t \Bigg) \\ &= 1 + O(\varepsilon n) + O\Bigg(\sum_{t=0}^{n-2} n^{n-t}(2\varepsilon)^{n-t} \Bigg) = 1 + O(\varepsilon n). \end{aligned}$$

Here we used the fact that there are $\leqslant \frac{n!}{t!} \leqslant n^{n-t}$ permutations σ with t fixed points, and made several uses of our assumption that $|\varepsilon n| \leqslant 1/4$.

Also

$$C^{-1} = (I + E' + (E')^2 + (E')^3 + \cdots)A^{-1},$$

where the infinite series converges because the entries of $(E')^j$ are (by an easy induction) bounded in absolute value by $n^{j-1}(2\varepsilon)^j \leqslant \frac{2\varepsilon}{2^{j-1}}$. Hence, $C^{-1} = (I + E'')A^{-1}$, where the entries of E'' are bounded in absolute value by 4ε. The second part now follows. $\square$

Proof of Proposition 2.3. First, note that $\det A = (1 - \xi^2)^k$ and recall that the inverse of $\left(\begin{smallmatrix} 1 & \xi \\ \xi & 1 \end{smallmatrix}\right)$ is $(1-\xi^2)^{-1}\left(\begin{smallmatrix} 1 & -\xi \\ -\xi & 1 \end{smallmatrix}\right)$. Therefore, it follows from (4.2) and Lemma 4.1 that

$$\begin{aligned} f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}(1-\xi^2)^{-\frac{k}{2}}(1 + O(\varepsilon n)) \times \\ \times \exp\left\{ -\frac{1}{2(1-\xi^2)} \sum_{j=1}^{k} (x_{2j-1}^2 + x_{2j}^2 - 2\xi x_{2j-1}x_{2j}) - \frac{1}{2}\sum_{j=2k+1}^{n} x_j^2 + \sum_{h,j=1}^{n} \varepsilon_{h,j} x_h x_j \right\}, \end{aligned}$$

where $|\varepsilon_{h,j}| \leqslant 8\varepsilon$ for every $1 \leqslant h, j \leqslant n$. If we write $\|\mathbf{x}_{2k}\| = (x_1^2 + \cdots + x_{2k}^2)^{1/2}$ for the Euclidean norm of the first $2k$ components of $\mathbf{x}$, and $\|\mathbf{x}^{n-2k}\| = (x_{2k+1}^2 + \cdots + x_n^2)^{1/2}$, and $\|\mathbf{x}\| = (x_1^2 + \cdots + x_n^2)^{1/2}$, then by the Cauchy–Schwarz inequality we obtain

$$\left| \sum_{h,j=1}^{n} \varepsilon_{h,j} x_h x_j \right| \leqslant 8\epsilon \left(\sum_{h=1}^{n} |x_h| \right)^2 \leqslant 8\varepsilon n \|\mathbf{x}\|^2 = 8\varepsilon n \left(\|\mathbf{x}_{2k}\|^2 + \|\mathbf{x}^{n-2k}\|^2 \right).$$

Thus we deduce that

$$\begin{aligned} f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}(1-\xi^2)^{-\frac{k}{2}}(1 + O(\varepsilon n)) \times \\ \times \exp\left\{ \frac{\xi}{1-\xi^2} \sum_{j=1}^{k} x_{2j-1}x_{2j} - \|\mathbf{x}_{2k}\|^2 \left(\frac{1}{2(1-\xi^2)} + O(\varepsilon n) \right) - \|\mathbf{x}^{n-2k}\|^2 \left(\frac{1}{2} + O(\varepsilon n) \right) \right\}. \end{aligned} \tag{4.3}$$

Invoking (2.16), we have that $\frac{1}{1-\xi^2} = 1 + O(1/\log^2 q)$. Also, by our assumption (2.1) and estimate (4.1) it follows that

$$\varepsilon n \ll \frac{1}{\log^2 q}.$$

Another application of Cauchy–Schwarz yields

$$\sum_{j=1}^{k} x_{2j-1}x_{2j} \leqslant \|\mathbf{x}_{2k}\|^2, \tag{4.4}$$

and therefore

$$\frac{\xi}{1-\xi^2}\sum_{j=1}^{k} x_{2j-1}x_{2j} = \xi\sum_{j=1}^{k} x_{2j-1}x_{2j} + O\left(\xi^3\|\mathbf{x}_{2k}\|^2\right)$$
$$= \xi\sum_{j=1}^{k} x_{2j-1}x_{2j} + O\left((\log q)^{-3}\|\mathbf{x}_{2k}\|^2\right).$$

Combining these estimates with (4.3) gives the first estimate in Proposition 2.3.

To obtain the crude estimate (2.18), we combine (2.16) and (4.4) to get

$$\xi\sum_{j=1}^{k} x_{2j-1}x_{2j} \ll \frac{\|\mathbf{x}_{2k}\|^2}{\log q}.$$

Inserting this bound into the first estimate of Proposition 2.3 completes the proof. □

We conclude this section with bounds on the tails of subvectors of $\mathbf{Z}$.

Lemma 4.2. *Assume* (2.1)*, and set* $\mathcal{Q} := \log q$*. Then*

$$\mathbb{P}\Big(\sum_{i=2k+1}^{n} Z_i^2 > 10n\log(n\mathcal{Q})\Big) \leqslant e^{-3n\log(n\mathcal{Q})+O(n)},$$
$$\mathbb{P}\Big(\sum_{i=1}^{2k} Z_i^2 > 10k\log(n\mathcal{Q})\Big) \leqslant e^{-3k\log(n\mathcal{Q})+O(k)}.$$

Proof of Lemma 4.2. Arguing as before, the density function of $(Z_i)_{2k+1\leqslant i\leqslant n}$ takes the form

$$\frac{(1+O(\varepsilon n))}{(2\pi)^{(n-2k)/2}}\exp\Big\{-\frac{1}{2}\|\mathbf{x}^{n-2k}\|^2(1+O(\varepsilon n))\Big\},$$

and since $\varepsilon n \ll \frac{1}{\log^2 q}$ we get

$$\mathbb{P}\Big(\sum_{i=2k+1}^{n} Z_i^2 > 10n\log(n\mathcal{Q})\Big) \ll \int_{\|\mathbf{x}^{n-2k}\|^2>10n\log(n\mathcal{Q})} \frac{e^{-\frac{1}{2}\|\mathbf{x}^{n-2k}\|^2(1+O(\frac{1}{\log^2 q}))}}{(2\pi)^{(n-2k)/2}}dx_{2k+1}...dx_n$$
$$\leqslant e^{-3n\log(n\mathcal{Q})}\left(\int_{-\infty}^{\infty}\frac{e^{-\frac{x^2}{6}}}{\sqrt{2\pi}}dx\right)^{n-2k} = e^{-3n\log(n\mathcal{Q})+O(n)}.$$

The proof of the second part of the lemma is exactly similar, using that (similarly as in Proposition 2.3) the density function of $(Z_i)_{1\leqslant i\leqslant 2k}$ takes the form

$$\frac{1}{(2\pi)^k}\exp\left\{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2\left(1+O\left(\frac{1}{\log q}\right)\right)+O\left(\frac{k}{(\log q)^2}\right)\right\}.$$

□

5. Auxiliary results on independent Gaussian random variables

As before

$$\mathbf{W} = (W_1, \ldots, W_n)$$

is a vector of independent standard Gaussian random variables, $h^{\pm}_{S,\alpha}$ are the smooth functions defined in (3.1), and $S^\flat$, $S^\sharp$ are the sets in (2.11), and (2.12) respectively. We start by proving a generalization of the upper bound (2.19).

Lemma 5.1. *Let* $1 \leqslant k \leqslant n/2$*, and let* $0 < \alpha \leqslant \frac{1}{5k^5 \log^2 n}$*. Then*

$$\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W}) \leqslant (1 + O(\sqrt{\alpha \log(1/\alpha) k^4 \log n}))\frac{(n-2k)!}{n!}.$$

We remark that the choice (2.15) of α satisfies the hypotheses of this lemma.

Proof of Lemma 5.1. By definition, we have $h^{+}_{S^\flat,\alpha}(\mathbf{W}) \leqslant g(-\sqrt{\alpha}) \ll e^{-1/\sqrt{\alpha}}$ unless $W_i - W_j \geqslant -2\sqrt{\alpha}$ for all pairs $(i,j) \in S^\flat$. Furthermore, in that case we still have $h^{+}_{S^\flat,\alpha}(\mathbf{W}) \leqslant 1$. So recalling our construction of $S^\flat$, if we define $\widetilde{W_i} := W_i + (2k+1-i)2\sqrt{\alpha}$ for $1 \leqslant i \leqslant 2k$, and define $\widetilde{W_i} := W_i$ for $2k+1 \leqslant i \leqslant n$, then

$$\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W}) \leqslant \mathbb{P}\left(\widetilde{W_1} > \widetilde{W_2} > \ldots > \widetilde{W_{2k}} > \max_{2k<i\leqslant n} \widetilde{W_i}\right) + O(e^{-1/\sqrt{\alpha}}).$$

Here our upper bound condition on α implies that

$$e^{-1/\sqrt{\alpha}} \ll \sqrt{\alpha \log(1/\alpha) k^4 \log n} \cdot n^{-2k} \leqslant \sqrt{\alpha \log(1/\alpha) k^4 \log n}\frac{(n-2k)!}{n!},$$

and so the "big Oh" term is acceptable for Lemma 5.1.

Next, by independence the probability here is simply

$$= \int_{x_1>x_2>\ldots>x_{2k}} \frac{e^{-\sum_{i=1}^{2k}(x_i-(2k+1-i)2\sqrt{\alpha})^2/2}}{(2\pi)^k}\left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}}dt\right)^{n-2k} dx_1 \ldots dx_{2k}$$
$$= \int_{x_1>x_2>\ldots>x_{2k}} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2+O(k^{3/2}\sqrt{\alpha}\|\mathbf{x}_{2k}\|+k^3\alpha)}}{(2\pi)^k}\left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}}dt\right)^{n-2k} dx_1 \ldots dx_{2k}.$$

The part of the integral where $\|\mathbf{x}_{2k}\|^2 > 10(k\log n + \log(1/\alpha))$ is

$$\ll \int_{\|\mathbf{x}_{2k}\|^2>10(k\log n+\log(1/\alpha))} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2+O(k^{3/2}\sqrt{\alpha}\|\mathbf{x}_{2k}\|)}}{(2\pi)^k} dx_1 \ldots dx_{2k} \leqslant e^{-3(k\log n+\log(1/\alpha))+O(k)},$$

which again is acceptable for Lemma 5.1. And the complementary part of the integral, where $\|\mathbf{x}_{2k}\|^2 \leqslant 10(k\log n + \log(1/\alpha))$, is

$$\leqslant e^{O\left(\sqrt{\alpha k^3(k\log n+\log(1/\alpha))}\right)} \int_{x_1>x_2>\ldots>x_{2k}} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2}}{(2\pi)^k} \left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}}dt\right)^{n-2k} dx_1 \ldots dx_{2k}.$$

Finally, the integral here is simply $\mathbb{P}\left(W_1 > W_2 > \ldots > W_{2k} > \max_{2k<i\leqslant n} W_i\right)$, which by symmetry is equal to $\frac{(n-2k)!}{n!}$. □

Next, we prove the following result which implies the asymptotic formulae in (2.20).

Lemma 5.2. *Let $1 \leqslant k \leqslant n/2$, and let $0 < \alpha \leqslant \frac{1}{5k^5\log^2 n}$. Then*

$$(1-O(n^2\sqrt{\alpha}))\frac{(n-2k)!}{n!} \leqslant \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W}) \leqslant \mathbb{E}h^+_{S^\sharp,\alpha}(\mathbf{W}) \leqslant (1+O(\sqrt{\alpha\log(1/\alpha)k^4\log n}))\frac{(n-2k)!}{n!}.$$

Proof of Lemma 5.2. For the upper bound, take

$$\begin{aligned} \widetilde{W}_{2i-1} &= W_{2i-1} + (k+1-i)2\sqrt{\alpha} \quad (1 \leqslant i \leqslant k), \\ \widetilde{W}_{2i} &= W_{2i} - (k+1-i)2\sqrt{\alpha} \quad (1 \leqslant i \leqslant k), \\ \widetilde{W}_i &= W_i \quad (2k+1 \leqslant i \leqslant n). \end{aligned}$$

Then, if $W_i - W_j \geqslant -2\sqrt{\alpha}$ for all $(i,j) \in S^\sharp$, we have

$$\widetilde{W}_1 > \widetilde{W}_3 > \cdots > \widetilde{W}_{2k-1} > \max_{2k+1\leqslant j\leqslant n} \widetilde{W}_j, \quad \widetilde{W}_2 < \widetilde{W}_4 < \cdots < \widetilde{W}_{2k} < \min_{2k+1\leqslant j\leqslant n} \widetilde{W}_j.$$

The upper bound then follows similarly as in Lemma 5.1.

For the lower bound, we note that if $(x_1,\ldots,x_n) \in R(S^\sharp)$ is such that $|x_i - x_j| \geqslant 2\sqrt{\alpha}$ for all $i \neq j$, then checking the definition (3.1) of $h^-_{S^\sharp,\alpha}$ we obtain

$$h^-_{S^\sharp,\alpha}(x_1,\ldots,x_n) \geqslant \prod_{(i,j)\in S^\sharp} g(\sqrt{\alpha}) \geqslant 1 - O(ne^{-1/\sqrt{\alpha}}).$$

Hence, if we temporarily let G denote the event that $|W_i - W_j| \geqslant 2\sqrt{\alpha}$ for all $i \neq j$, then we see that $\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})$ is at least

$$\begin{aligned} (1 - O(ne^{-1/\sqrt{\alpha}}))\mathbb{P}\big(W_1 > W_3 > \ldots > W_{2k-1} > \max_{2k<i\leqslant n} W_i \\ \geqslant \min_{2k<i\leqslant n} W_i > W_{2k} > \cdots > W_4 > W_2, \text{ and } G\big). \end{aligned} \tag{5.1}$$

However, by symmetry the probability here is equal to $\frac{(n-2k)!}{n!}\mathbb{P}(G)$, and by the union bound we have

$$\mathbb{P}(G) \geqslant 1 - \sum_{i\neq j} \mathbb{P}(|W_i - W_j| \leqslant 2\sqrt{\alpha}) = 1 - O(n^2\sqrt{\alpha}).$$ □

We end this section with the following large deviation estimate, which is analogous to Lemma 4.2, but is even easier to prove because of independence.

Lemma 5.3. *Let $W_1, \ldots, W_k$ be independent standard Gaussian random variables. For all $V \geqslant 0$ we have*

$$\mathbb{P}\Big(\sum_{i=1}^{k} W_i^2 > V\Big) \leqslant e^{-V/3+O(k)},$$

Proof of Lemma 5.3. We have

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^{k} W_i^2 > V\Big) &= \int_{\sum_{i=1}^{k} x_i^2 > V} \frac{e^{-\frac{1}{2}\sum_{i=1}^{k} x_i^2}}{(2\pi)^{k/2}} dx_1 ... dx_k \\ &\leqslant e^{-V/3}\left(\int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{6}}}{\sqrt{2\pi}} dx\right)^k = e^{-V/3+O(k)}.\end{aligned}$$

□

6. Factoring out the bias: Proofs of Propositions 2.4 and 2.5

We let $(a_1, \ldots, a_n)$ be the tuple of residues satisfying the conclusion of Lemma 2.2, and $\mathbf{Z} = (Z_i)_{1\leqslant i\leqslant n}$ be the corresponding vector of jointly Gaussian random variables of mean 0, variance 1, and covariances given by (2.9). As before, we take $\alpha = 1/(n^5 \log^6 q)$ as in (2.15), and also let $h^{\pm}_{S,\alpha}$ be the smooth functions defined in (3.1), $\mathbf{W} = (W_1, \ldots, W_n)$ be a vector of independent standard Gaussian random variables, and $S^\flat, S^\sharp$ be the sets in (2.11), and (2.12) respectively. To shorten the notation we put $\mathcal{Q} = \log q$.

Proof of Proposition 2.4. We first suppose that $k \log(n/k) > 1000 \log\log q$. (The case where $k \log(n/k) \leqslant 1000 \log\log q$ is a bit more delicate because then the bias factor $\exp\left(-\frac{0.6k\log(n/k)}{\log q}\right)$ we are aiming for is only slightly smaller than 1, so that will be dealt with later.) Let us introduce notation for the following three "good" events:

$$\mathcal{A} := \left\{ \sum_{i=2k+1}^{n} Z_i^2 \leqslant 10n\log(n\mathcal{Q}),\ \sum_{i=1}^{2k} Z_i^2 \leqslant 10k\log(n\mathcal{Q}) \right\}, \tag{6.1}$$

$$\mathcal{C} := \left\{ \max_{2k<i\leqslant n} Z_i > \sqrt{\log(n/k)} \right\}, \qquad \mathcal{U} := \{\min(Z_1, ..., Z_{2k}) \geqslant \max_{2k<i\leqslant n} Z_i - 2\sqrt{\alpha}\}. \tag{6.2}$$

Firstly, Lemma 4.2 and the definition of $h^+_{S^\flat,\alpha}$ (in particular the fact that $0 \leqslant h^+_{S^\flat,\alpha} \leqslant 1$) imply that

$$\begin{aligned}
|\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z}) - \mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}}| &\leqslant (1-\mathbb{P}(\mathcal{A})) + \mathbb{E}\big(h^+_{S^\flat,\alpha}(\mathbf{Z})\,(1-\mathbf{1}_{\mathcal{U}})\big) \\
&\ll e^{-3k\log(n\mathcal{Q})+O(k)} + g(-\sqrt{\alpha}) \\
&\ll e^{-3k\log(n\mathcal{Q})+O(k)} + e^{-(n\log q)^{5/2}} \\
&\ll e^{-3k\log(n\mathcal{Q})+O(k)} \ll \frac{(n-2k)!}{n!}e^{-k\log(n\mathcal{Q})}. \qquad (6.3)
\end{aligned}$$

Next, introduce the "truncated" set

$$S' = \{(i,i+1) : 1 \leqslant i \leqslant 2k-1\},$$

considered as a subset of $\{1,\dots,2k\}^2$, the associated set $R(S') \subset \mathbb{R}^{2k}$ and function $h^+_{S',\alpha} : \mathbb{R}^{2k} \to \mathbb{R}$. Using the crude estimate from Proposition 2.3 we have

$$\begin{aligned}
\mathbb{E}h^+_{S^\flat,\alpha}(Z)\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\{\mathcal{C}\text{ fails}\}} \leqslant\ & e^{O(\frac{k}{\log^2 q})} \underset{\|\mathbf{x}_{2k}\|^2 \leqslant 10k\log(n\mathcal{Q})}{\int\cdots\int} h^+_{S',\alpha}(x_1,\dots,x_{2k})\frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}(1+O(\frac{1}{\log q}))}}{(2\pi)^k}dx_1\cdots dx_{2k} \\
& \times \underset{\substack{\max(x_{2k+1},\dots,x_n)\leqslant\sqrt{\log(n/k)},\\ \|\mathbf{x}^{n-2k}\|^2\leqslant 10n\log(n\mathcal{Q})}}{\int\cdots\int} \frac{e^{-\frac{\|\mathbf{x}^{n-2k}\|^2}{2}(1+O(\frac{1}{n\log^{10}q}))}}{(2\pi)^{(n-2k)/2}}dx_{2k+1}\dots dx_n. \qquad (6.4)
\end{aligned}$$

Now the integral on the first line here is

$$\leqslant e^{O(\frac{k\log(n\mathcal{Q})}{\log q})}\int h^+_{S',\alpha}(x_1,\dots,x_{2k})\frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}}}{(2\pi)^k}dx_1\cdots dx_{2k} \ll e^{O(\frac{k\log(n\mathcal{Q})}{\log q})}\frac{1}{(2k)!} = \frac{e^{O(k)}}{k^{2k}},$$

where the second inequality follows from Lemma 5.1 (with n replaced by $2k$), and the final equality follows from Stirling's formula. Meanwhile, the integral on the second line in (6.4) is

$$\ll \left(\int_{-\infty}^{\sqrt{\log(n/k)}}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx\right)^{n-2k} = \left(1-\Theta\left(\frac{e^{-\log(n/k)/2}}{\sqrt{\log(n/k)}}\right)\right)^{n-2k} = \exp\left\{-\Theta\left(\sqrt{\frac{nk}{\log(n/k)}}\right)\right\}.$$

(Notice here that $n/k \geqslant A$ is large, under our hypotheses.) So putting things together, we have shown that

$$\begin{aligned}\mathbb{E}\left\{h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\{\mathcal{C}\text{ fails}\}}\right\} \leqslant \frac{e^{O(k)}}{k^{2k}}e^{-\Theta\left(\sqrt{\frac{nk}{\log(n/k)}}\right)} &= \frac{1}{n^{2k}}e^{O(k\log(n/k))-\Theta\left(k\sqrt{\frac{n/k}{\log(n/k)}}\right)} \\ &\leqslant \frac{1}{n^{2k}}e^{-1000k\log(n/k)} \\ &\leqslant \frac{(n-2k)!}{n!}e^{-1000k\log(n/k)}.\end{aligned} \tag{6.5}$$

On the other hand, the first part of Proposition 2.3 implies that on those tuples $(x_1, ..., x_n)$ corresponding to all three events $\mathcal{A}, \mathcal{C}, \mathcal{U}$, the density function $f(x_1, \ldots, x_n)$ satisfies

$$\begin{aligned}f(\mathbf{x}) &\leqslant \frac{e^{\xi k\log(n/k)+O(\frac{k}{\log^2 q})}}{(2\pi)^{n/2}}\exp\left\{-\frac{\|\mathbf{x}_{2k}\|^2}{2}\Big(1+O\Big(\frac{1}{\log^2 q}\Big)\Big)-\frac{\|\mathbf{x}^{n-2k}\|^2}{2}\Big(1+O\Big(\frac{1}{n\log^{10} q}\Big)\Big)\right\} \\ &= \frac{e^{\xi k\log(n/k)+O\left(\frac{k\log(n\mathcal{Q})}{\log^2 q}\right)}}{(2\pi)^{n/2}}\exp\left\{-\frac{\|\mathbf{x}\|^2}{2}\right\},\end{aligned}$$

where $\xi \sim -\frac{\log 2}{\log q}$ (recall (2.16)). The crucial thing to notice here is the emergence of the bias term $e^{\xi k\log(n/k)}$, which came from the non-trivial correlations of size ξ amongst pairs $(Z_{2i-1}, Z_{2i})_{1\leqslant i\leqslant k}$, together with the fact that we have arranged to have $Z_1, ..., Z_{2k} \geqslant \sqrt{\log(n/k)} - 2\sqrt{\alpha}$. Using this upper bound, we get

$$\begin{aligned}\mathbb{E}\left\{h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{C}\cap\mathcal{U}}\right\} &\leqslant e^{\xi k\log(n/k)+O(\frac{k\log(n\mathcal{Q})}{\log^2 q})}\int h^+_{S^\flat,\alpha}(x_1,\ldots,x_n)\frac{e^{-\frac{\|\mathbf{x}\|^2}{2}}}{(2\pi)^{n/2}}dx_1...dx_n \\ &= e^{\xi k\log(n/k)+O(\frac{k\log(n\mathcal{Q})}{\log^2 q})}\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W}) \leqslant e^{\xi k\log(n/k)+O(\frac{k}{\log q})}\frac{(n-2k)!}{n!},\end{aligned} \tag{6.6}$$

where the second inequality follows from (2.1) and Lemma 2.6.

Again recall (2.16). Hence, for q sufficiently large, and A sufficiently large, the "big Oh" term in the exponent in (6.6) is smaller than $0.01|\xi k\log(n/k)|$ and consequently

$$\xi k\log(n/k) + O\left(\frac{k}{\log q}\right) \leqslant -0.65\frac{k\log(n/k)}{\log q}.$$

Adding this together with (6.3) and (6.5), we obtain that $\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z})$ is

$$\begin{aligned}&\leqslant \left(\exp\left(-\frac{0.65k\log(n/k)}{\log q}\right)+O(e^{-k\log(n\mathcal{Q})})+e^{-1000k\log(n/k)}\right)\frac{(n-2k)!}{n!} \\ &\leqslant \left(\exp\left(-\frac{0.05k\log(n/k)}{\log q}\right)+O(\frac{1}{\log^k q})+e^{-999k\log(n/k)}\right)\exp\left(-\frac{0.6k\log(n/k)}{\log q}\right)\frac{(n-2k)!}{n!} \\ &\leqslant \exp\left(-\frac{0.6k\log(n/k)}{\log q}\right)\frac{(n-2k)!}{n!},\end{aligned}$$

since $k\log(n/k) > 1000\log\log q$. (It is precisely here that we use this condition, to control the contribution from the third term at the final step.) This proves the required bound in Proposition 2.4 in the case where $k\log(n/k) > 1000\log\log q$.

Next we turn to the other case where $k\log(n/k) \leqslant 1000\log\log q$. Notice that in this case, in the definition of the event $\mathcal{A}$ we have $10k\log(n\mathcal{Q}) = 10k(\log k + \log(n/k) + \log\mathcal{Q}) \ll (\log\log q)^2$, and so we get $\|\mathbf{x}_{2k}\|^2 \ll (\log\log q)^2$. Write

$$\mathcal{A}' := \left\{ \sum_{i=2k+1}^{n} W_i^2 \leqslant 10n\log(n\mathcal{Q}), \ \ \sum_{i=1}^{2k} W_i^2 \leqslant 10k\log(n\mathcal{Q}) \right\},$$
$$\mathcal{C}' := \left\{ \max_{2k<i\leqslant n} W_i > \sqrt{\log(n/k)} \right\}, \qquad \mathcal{U}' := \{\min(W_1, ..., W_{2k}) \geqslant \max_{2k<i\leqslant n} W_i - 2\sqrt{\alpha}\}.$$

the events corresponding to $\mathcal{A},\mathcal{C},\mathcal{U}$ but with W_i replacing Z_i. Then the above calculations in fact imply that

$$\begin{aligned}\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\mathcal{C}} &\leqslant e^{\xi k\log(n/k)+O(\frac{k}{\log q})}\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\mathcal{C}'} \\ &= \left(1+\xi k\log(n/k) + O\left(\frac{k}{\log q}\right)\right)\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\mathcal{C}'},\end{aligned}$$

and that (using the crude density estimate from Proposition 2.3, as in (6.4))

$$\begin{aligned}&|\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\{\mathcal{C}\text{ fails}\}} - \mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\{\mathcal{C}'\text{ fails}\}}| \\ &\qquad \ll \int_{\mathcal{A}\cap\mathcal{U}\cap\{\mathcal{C}\text{ fails}\}} h^{+}_{S^\flat,\alpha}(x_1,\ldots,x_n)\frac{e^{-\frac{||\mathbf{x}||^2}{2}}}{(2\pi)^{n/2}}\left(\frac{k}{\log^2 q} + \frac{||\mathbf{x}_{2k}||^2}{\log q}\right)dx_1...dx_n.\end{aligned}$$

The part of this integral where $\|\mathbf{x}_{2k}\|^2 \leqslant 10k\log k$ contributes $\ll \frac{k\log k}{\log q}\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\{\mathcal{C}'\text{ fails}\}}$, which is $\ll \frac{1}{\log q}\frac{(n-2k)!}{n!}$ by the calculations leading to (6.5). Meanwhile, another short calculation shows the other part of the integral contributes

$$\begin{aligned}&\ll \frac{1}{\log q}\int_{\|\mathbf{x}_{2k}\|^2>10k\log k}\frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}}}{(2\pi)^k}\|\mathbf{x}_{2k}\|^2 dx_1...dx_{2k}\cdot\mathbb{P}(\max_{2k<i\leqslant n} W_i \leqslant \sqrt{\log(n/k)}) \\ &\ll \frac{1}{\log q}e^{-3k\log k+O(k)}e^{-\Theta\left(\sqrt{\frac{nk}{\log(n/k)}}\right)} \ll \frac{1}{\log q}\frac{(n-2k)!}{n!}.\end{aligned}$$

The calculations leading to (6.5) further imply that

$$\xi k\log(n/k)\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\{\mathcal{C}'\text{ fails}\}} \ll \frac{k\log(n/k)}{\log q}\mathbb{E}h^{+}_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'\cap\{\mathcal{C}'\text{ fails}\}} \ll \frac{1}{\log q}\frac{(n-2k)!}{n!}.$$

So we can collect together the above computations, along with (6.3), in the form

$$\begin{aligned}\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z}) &= \mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\mathcal{C}} + \mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}\cap\{\mathcal{C}\text{ fails}\}} + (\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z}) - \mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{U}})\\ &\leqslant \left(1+\xi k\log(n/k) + O\left(\frac{k}{\log q}\right)\right)\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'} + O\left(\frac{1}{\log q}\frac{(n-2k)!}{n!}\right).\end{aligned}$$

And Lemma 2.6 implies that $\mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{U}'} \leqslant \mathbb{E}h^+_{S^\flat,\alpha}(\mathbf{W}) \leqslant (1+O(1/\log^2 q))\frac{(n-2k)!}{n!}$, which suffices to establish the proposition in the case where $k\log(n/k) \leqslant 1000\log\log q$.

□

Proof of Proposition 2.5. The proof is similar to that of Proposition 2.4. In addition to the events $\mathcal{A},\mathcal{C}$ defined in (6.1) and (6.2), we introduce the following:

$$\mathcal{D} := \left\{\min_{2k<i\leqslant n} Z_i < -\sqrt{\log(n/k)}\right\},$$

$$\mathcal{V} := \left\{Z_1, Z_3, ..., Z_{2k-1} \geqslant \max_{2k<i\leqslant n} Z_i - 2\sqrt{\alpha}, \quad\text{and}\quad Z_2, Z_4, ..., Z_{2k} \leqslant \min_{2k<i\leqslant n} Z_i + 2\sqrt{\alpha}\right\}.$$

As in the proof of Proposition 2.4, on those tuples $(x_1, ..., x_n)$ corresponding to all the events $\mathcal{A},\mathcal{C},\mathcal{D},\mathcal{V}$, the density function $f(x_1,\dots,x_n)$ satisfies

$$f(\mathbf{x}) \geqslant \frac{e^{-\xi k\log(n/k) + O\left(\frac{k\log(n\mathcal{Q})}{\log^2 q}\right)}}{(2\pi)^{n/2}}\exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right).$$

Again, the key thing to notice here is the large bias factor $e^{-\xi k\log(n/k)}$, which emerged because we arranged to have

$$\begin{aligned}\min(Z_1, Z_3, ..., Z_{2k-1}) &\geqslant \sqrt{\log(n/k)} - 2\sqrt{\alpha}, \quad\text{and}\\ \max(Z_2, Z_4, ..., Z_{2k}) &\leqslant -(\sqrt{\log(n/k)} - 2\sqrt{\alpha}).\end{aligned}$$

So, we have that

$$\begin{aligned}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z}) &\geqslant \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{V}} = \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{C}\cap\mathcal{D}\cap\mathcal{V}} + \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{V}\cap\{\mathcal{C}\text{ or }\mathcal{D}\text{ fails}\}}\\ &\geqslant e^{-\xi k\log(n/k) + O\left(\frac{k\log(n\mathcal{Q})}{\log^2 q}\right)}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{C}'\cap\mathcal{D}'\cap\mathcal{V}'} + \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{V}\cap\{\mathcal{C}\text{ or }\mathcal{D}\text{ fails}\}},\end{aligned}$$

where the 'primed' events (e.g., $\mathcal{A}'$) are the corresponding events with each Z_i replaced by W_i. As before, for A and q large enough, the "big Oh" term in the exponent is negligible and we have $e^{-\xi k\log(n/k) + O\left(\frac{k\log(n\mathcal{Q})}{\log^2 q}\right)} \geqslant e^{0.65k\frac{\log(n/k)}{\log q}}$, and so

$$\begin{aligned}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z}) &\geqslant e^{0.65k\frac{\log(n/k)}{\log q}}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{V}'}\\ &\quad + \mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\{\mathcal{C}\text{ or }\mathcal{D}\text{ fails}\}\cap\mathcal{V}} - e^{0.65k\frac{\log(n/k)}{\log q}}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\{\mathcal{C}'\text{ or }\mathcal{D}'\text{ fails}\}\cap\mathcal{V}'}.\end{aligned} \tag{6.7}$$

Furthermore, Lemmas 5.3 and 2.6 together imply, with our choice (2.15), that

$$\begin{aligned}\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{V}'} &\geqslant (1-O(n^2\sqrt{\alpha}))\frac{(n-2k)!}{n!}+O(e^{-3k\log(n\mathcal{Q})+O(k)})\\ &=\left(1-O\left(\frac{1}{\log^{5/2}q}\right)\right)\frac{(n-2k)!}{n!}.\end{aligned}$$

We can also mimic the calculations leading to (6.5), this time using the second estimate in Lemma 2.6 to bound $\int h^-_{S^\sharp,\alpha}(x_1,\dots,x_{2k})\frac{e^{-\frac12\|\mathbf{x}_{2k}\|^2}}{(2\pi)^k}dx_1\cdots dx_{2k}$, and obtain that

$$\mathbb{E}\left\{h^-_{S^\sharp,\alpha}(\mathbf{W})\mathbf{1}_{\mathcal{A}'\cap\mathcal{V}'\cap\{\mathcal{C}' \text{ or } \mathcal{D}' \text{ fails}\}}\right\}\leqslant\frac{1}{n^{2k}}e^{-1000k\log(n/k)}\leqslant\frac{(n-2k)!}{n!}e^{-1000k\log(n/k)},$$

and the same for $\mathbb{E}\left\{h^-_{S^\sharp,\alpha}(\mathbf{Z})\mathbf{1}_{\mathcal{A}\cap\mathcal{V}\cap\{\mathcal{C} \text{ or } \mathcal{D} \text{ fails}\}}\right\}$. Hence when $k\log(n/k)>1000\log\log q$, the final two terms in (6.7) make a negligible contribution, and the desired lower bound for $\mathbb{E}h^-_{S^\sharp,\alpha}(\mathbf{Z})$ follows. In the remaining case where $k\log(n/k)\leqslant 1000\log\log q$, we can again duplicate our previous approach and show the *difference* of the final two terms in (6.7) is $\ll\frac{1}{\log q}\frac{(n-2k)!}{n!}$, which implies the result. □

Acknowledgements. We would like to thank the anonymous referee for carefully reading the paper, and for thoughtful remarks and suggestions.

References

[1] V. Y. Bentkus, *Large deviations in Banach spaces.* Theory Probab. Appl. **31** (1986), no. 4, 627-632.

[2] A. Feuerverger and G. Martin, *Biases in the Shanks-Rényi prime number race.* Experiment. Math. **9** (2000), no. 4, 535-570.

[3] D. Fiorilli, *Highly biased prime number races.* Algebra Number Theory **8** (2014), no. 7, 1733-1767.

[4] D. Fiorilli and G. Martin, *Inequities in the Shanks-Rényi Prime Number Race: An asymptotic formula for the densities.* J. Reine Angew. Math. **676** (2013), 121-212.

[5] K. Ford and S. Konyagin, *Chebyshev's conjecture and the prime number race.* Modern Problems of Number Theory and its Applications; Topical Problems Part II (Tula, Russia, 2001).

[6] K. Ford and S. Konyagin, *The prime number race and zeros of L-functions off the critical line. II.* Proceedings of the Session in Analytic Number Theory and Diophantine Equations, 40 pp., Bonner Math. Schriften, 360, Univ. Bonn, Bonn, 2003.

[7] A. Granville and G. Martin, *Prime number races.* Amer. Math. Monthly **113** (2006), no. 1, 1-33.

[8] A. J. Harper and Y. Lamzouri, *Orderings of weakly correlated random variables, and prime number races with many contestants.* Probab. Theory Related Fields **170** (2018), no. 3-4, 961-1010.

[9] C. Hooley, *On the Barban–Davenport–Halberstam theorem: VII*, J. London Math. Soc. (2) **16** (1977), 1–8.

[10] Y. Lamzouri, *The Shanks-Rényi prime number race with many contestants.* Math. Res. Lett. **19** (2012), no. 03, 649-666.

[11] Y. Lamzouri, *Prime number races with three or more competitors*, Math. Annalen **356** (2013), 1117–1162.

[12] G. Martin and J. Scarfy, *Comparative prime number theory: A survey.* 37 pages. `arXiv:1202.3408`

[13] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I: Classical Theory.* First edition, published by Cambridge University Press, 2007.
[14] M. Rubinstein and P. Sarnak, *Chebyshev's bias.* Experiment. Math. **3** (1994), no. 3, 173-197.

KF: Department of Mathematics, University of Illinois at Urbana-Champaign, 1409 West Green Street, Urbana, IL 61801, USA

E-mail address: ford@math.uiuc.edu

AJH: Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, England

E-mail address: A.Harper@warwick.ac.uk

YL: Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON, M3J1P3 Canada

E-mail address: lamzouri@mathstat.yorku.ca